The rule-based insensitivity effect: a systematic review

Ama Kissi[1], Colin Harte[1], Sean Hughes[1], Jan De Houwer[1], Geert Crombez[1]

[1] Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, East-Flanders, Belgium

Corresponding Author:
Ama Kissi[1]
Henri Dunantlaan 2, 9000 Gent, Belgium.
Email address: Ama.Kissi@UGent.be

Abstract

Background. Adherence to inaccurate rules has been viewed as a characteristic of human rule-following (i.e., the rule-based insensitivity effect; RBIE) and has been thought to be exacerbated in individuals suffering from clinical conditions. This review intended to systematically examine these claims in adult populations.

Methodology. We screened 1464 records which resulted in 21 studies that were deemed eligible for inclusion. Each of these studies was examined to determine: (1) if there is evidence for the RBIE in adults and (2) if this effect is larger in those suffering from psychological problems compared to their non-suffering counterparts. In addition, we investigated how (3) different operationalizations of the RBIE, and (4) the external validity and risks of bias of the experimental work investigating this effect, might influence the conclusions that can be drawn from the current systematic review.

Results. (1) Out of the 20 studies that were relevant for examining if evidence exists for the RBIE in adults, only 11 were eligible for vote counting. Results showed that after the contingency change, the rule groups were more inclined to demonstrate behavior that was reinforced before the change, compared to their non-instructed counterparts. Critically, however, none of these studies examined if their no-instructions group was an adequate comparison group. As a result, this made it difficult to determine whether the effects that were observed in the rules groups could be attributed to the rules or instructions that were manipulated in those experiments. (2) The single study that was relevant for examining if adults suffering from psychological problems demonstrated larger levels of the RBIE, compared to their non-clinical counterparts, was not eligible for vote counting. As a result, no conclusions could be drawn about the extent to which psychological problems moderated the RBIE in that study. (3) Similar procedures and tasks have been used to examine the RBIE, but their precise parameters differ across studies; and (4) most studies report insufficient information to evaluate all relevant aspects affecting their external validity and risks of bias.

Conclusions. Despite the widespread appeal that the RBIE has enjoyed, this systematic review indicates that, at present, only preliminary evidence exists for the idea that adults demonstrate the RBIE and no evidence is available to assume that psychological problems exacerbate the RBIE in adults.

The systematic review was registered in PROSPERO (CRD42018088210).

Introduction

Rules[1] constitute a set of statements that can govern behavior in various domains such as personal, professional, social, and legal contexts. In most cases adherence to rules like "eat healthily if you want to live long," "do not offend your boss," "do not gossip about your friends," and "do not drink and drive" is beneficial, in so far as doing so allows the individual to more readily obtain positive consequences (e.g., a long life, job certainty) or avoid negative ones (e.g., losing your friends, getting a fine). Yet despite the consequences of rule-following, rules can also continue to exert control over behavior even when they are no longer accurate. Within the behavioral-analytic literature, this pattern of behavior has been referred to as the "*rule-based insensitivity effect*" (RBIE) and has been defined as "an insensitivity of behavior to other contingencies [2] due to rule-following" (see Kissi, Hughes, De Schryver, De Houwer, & Crombez, 2018, p. 1).

To illustrate this effect more clearly, consider the following example. Imagine participants are asked to complete a learning task and are assigned to one of two groups: an instructions or no-instructions group. In both groups, they can initially earn points if they press the spacebar rapidly in the presence of a green square. Before starting the task, the instructions group is accurately informed about the contingencies operating in the task (i.e., that pressing the spacebar *rapidly* will cause them to earn more points). The no-instructions group, however, is not informed about these contingencies and thus has to figure out how to earn points via trial-and-error. About half way through the task, the task-contingencies are changed so that participants now have to press the spacebar *slowly* in order to earn points. Under such circumstances, it would be assumed that there is evidence for the RBIE if participants who were initially provided with accurate instructions, earned fewer points *after the task-contingency change* compared to those that did not receive such instructions (see Kissi et al., 2018 and LeFrancois et al., 1988 for similar procedures).

Over the past decades, a number of studies have empirically examined the RBIE in the laboratory (e.g., Donadeli & Strapasson, 2015; Joyce & Chase, 1990; Miller, Hirst, Kaplan, DiGennaro Reed, & Reed, 2014; Ninness & Ninness, 1998). Elsewhere, applied researchers and clinical psychologists have appealed to this effect when attempting to understand and treat psychological suffering. For instance, it has been argued that the RBIE is at the core of various problems such as addiction, depression, and personality disorders (Baruch, Kanter, Busch, Richardson, & Barnes-Holmes, 2007; Blackledge & Drake, 2013; Hayes & Gifford, 1997; McAuliffe, Hughes, & Barnes-Holmes, 2014; Törneke, Luciano, & Salas, 2008; Törneke, 2010). The idea here is that psychological problems are – amongst other things – the consequence of

[1] Within the behavioral-analytic literature terms such as instructions and rules are often used interchangeably. Yet it is important to note that they are descriptive and not functional-analytical terms, given that they did not emerge from inductive, functional-analytic research. As such, in the current manuscript we will use them interchangeably as a way to orient the reader toward a specific class of verbal stimuli.

[2] These contingencies can refer to other contingencies in the environment as well as those specified by a rule.

adherence to rules that reduce one's ability to persist or adapt to what is required in a given situation (Blackledge & Drake, 2013).

Nevertheless, and despite the attention that rules and the RBIE have received, there is currently no systematic review available of the experimental work examining this effect. This is unfortunate, given that such a review is essential to draw general conclusions about the RBIE which can inform future research and clinical practice. Towards this end, we systematically reviewed the RBIE literature to examine if: (1) there is sufficient empirical support for this effect in adults, and (2) adults suffering from psychological problems display larger levels of this effect compared to those that do not suffer from these problems. We also investigated how (3) different operationalizations of the RBIE, and (4) the external validity and risks of bias of the experimental work investigating this effect, might influence the conclusions that can be drawn from the current systematic review.

Survey methodology

Protocol and Registration

The review protocol was designed in line with the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009) and registered in PROSPERO (CRD42018088210).

Information Sources and Search Strategy

To identify as many relevant records as possible, multiple electronic databases were searched (i.e., "Web of Science", "PsychINFO", "PsychArticles", and "PubMed [Medline]") using the search terms: *"rule governed behavior", "rule-governed behavior", "rule governed behaviour", "rule-governed behaviour", "verbal regulation", "instructional control", "verbal rule", "instructed behavior", "instructed behaviour", "instructed learning", "instruction following", "instruction-following", "rule following", and "rule-following."* These search terms were iteratively developed with experts on systematic reviews and rule-governed behavior, and were subsequently presented to other experts on systematic reviews and rule-governed behavior who were not associated with the project. All searches were conducted on 4/10/2017 by the first author (i.e., Ama Kissi) and yielded 1459 records. Five novel records were additionally retrieved by contacting experts in the field, which resulted in a final set of 1464 records that were assessed for eligibility.

Eligibility Criteria

There were several general criteria that a record had to meet before being included in the current review: (1) it had to be a peer-reviewed journal article, (2) it had to be written in English, (3) it had to include a study that examined the RBIE by first asking participants to follow socially –or self-generated rules that initially corresponded with a set of contingencies but then became inaccurate after a contingency change, and (4) this study had to have an overall sample

age of at least 18 years, (5) and at least 10 participants within each experimental group (see Van Ryckeghem, Van Damme, Eccleston, & Crombez, 2018 for similar eligibility criteria).

Furthermore, depending on the research objective under scrutiny, the individual studies reported in these records had to meet an additional number of criteria to be deemed eligible for inclusion. For instance, when addressing our first research question ("*Is there evidence for the rule-based insensitivity effect in adults?"*), we only included studies that did not focus upon individuals with clinical problems. That is, only studies which used convenience samples (e.g., students), samples taken from the general population, or those that were not diagnosed with clinical problems, or reported sub-clinical problems were included. Studies were deemed eligible for answering our second research question (*"Do adults suffering from psychological problems display a larger RBIE compared to their non-clinical counterparts?"*), if they used the following samples: individuals diagnosed with psychological problems (clinical group) or those who scored high on instruments measuring psychological problems but were not formally diagnosed with a clinical problem (sub-clinical group), *and* a comparison group consisting of individuals that did not suffer from the above problems or were recruited via convenience sampling.

Study Selection Process

Out of the 1464 records that were assessed for eligibility, 1446 were excluded because they were not published in English (n = 123), were not peer-reviewed journal articles (n = 207) (e.g., book chapters, dissertations, or conference papers) or dealt with a topic that did not meet our inclusion criteria (n = 1044). Three journal articles were, furthermore, omitted because they did not provide sufficient information to assess their eligibility. An additional 69 journal articles were excluded that were on the RBIE but were non-experimental (n = 6), relied on non-adult samples (n = 14), used samples with less than 10 participants per experimental condition (n = 41), or did not include a contingency change or manipulate accurate rules (n = 8). This resulted in a remaining total of 18 records consisting of 22 individual studies. One of these studies was subsequently omitted because it did not have at least 10 participants within each experimental group. As such, 21 studies were finally included in the systematic review. The eligibility of all studies were independently assessed by the first two reviewers (i.e., Ama Kissi and Colin Harte) initial agreement = 99% [kappa = .98], agreement after discussion = 100% [kappa = 1.00]). See Fig. 1 for the flow diagram of the study selection process.

----------INSERT FIGURE 1 HERE----------

Qualitative Synthesis: Coding Procedure and Items

Certain characteristics of each of the 21 studies were independently coded by the first two reviewers (i.e., Ama Kissi and Colin Harte) (initial inter-reviewer agreement = 96%, inter-reviewer agreement after discussion = 100%). These characteristics involved the source, study, task, and sample characteristics. The source characteristics entailed the year in which the first author published the study and the country where s/he worked in when the paper was published.

The study characteristics referred to the type of task, experimental design, procedure, and analytic method that were used to examine the RBIE. Furthermore, the task characteristics entailed whether a study reported the exact instructions or rules that were used, how these instructions or rules were delivered (orally versus written) or generated (self [i.e., by the rule-follower]-versus socially [i.e., by another person than the rule-follower]), the reinforcement schedules that were used, the required behavioral responses, the type of consequential stimuli that were used, whether the contingency change was (un)signaled, whether a description was provided of who the experimenter was, and whether the experimenter was present. Finally, the sample characteristics that were evaluated were the size and mean age of the sample, the ratio of males:females, and whether the sample was selected (i.e., from either a healthy, clinical or sub-clinical population, or the general population) or non-selected (i.e., a convenience sample). These characteristics were evaluated for each experimental group.

Quantitative Synthesis: Vote Counting

To synthesize the quantitative results of the included studies, we used the vote-counting method. This method was chosen because not all studies reported effect sizes or information that could be used to calculate such estimates. According to the Cochrane Collaboration guidelines for systematic reviews, the best way to use the vote-counting method is by assessing whether the results of the empirical studies fall into one of two categories: "*positive*" or "*negative*" effects (see Deeks, Higgins & Altman, 2008). *Positive* effects refer to results that are in favor of the predicted relationship between the independent and dependent variable(s), whereas *negative* effects refer to outcomes that are in the opposite direction of what is expected. We only judged (or voted) whether a study had positive or negative effects if it included a comparison group (i.e., a no-instructions group). That is, a group that received the same treatment as the rules group but was not asked to follow the instructions or rules that these groups had to follow. We applied this restriction because we argued that such a comparison group is necessary if a study wishes to draw conclusions about the extent to which certain rules or instructions are responsible for the observed effects. In doing so, performances in the comparison group would serve as a baseline of how people behave in the absence of these types of rules or instructions. As such, if a study did not include such a comparison group, we argued that its effects were *unclear* (i.e., there was insufficient information to cast votes).

The outcome data that were preferably used to cast votes were measures of the central tendency (e.g., mean, mode, or median) of participants' responses, during all blocks after the contingency change. If a study, however, did not report participants' performances during all blocks following the contingency change, but only during a fraction of the trials after this change, we limited our analysis to that data. In the unfortunate event that no data was provided that could be used to draw conclusions about the central tendency of participants' responding after the contingency change, we relied on the conclusions that the authors formulated themselves (Cerutti, 1991; Torgrud, Holborn, & Zak, 2006 [Experiments 1 and 2]). Finally, in all of the above cases, if there were multiple contingency changes we only considered participants'

responding after the *first* change. This was, specifically, done to prevent carry-over effects from influencing the interpretation of the results.

All votes were independently cast by the first two reviewers (i.e., Ama Kissi and Colin Harte) in the following manner (inter-reviewer agreement = 100%, kappa = 1.00). For the first research question ("*Is there evidence for the rule-based insensitivity effect in adults"*), study results were considered *positive* if evidence was found for the RBIE. That is, if participants did not adapt to a novel task-contingency or rule (i.e., if their behavior was still in line with the self-generated or socially-provided rule that was in place before the contingency change). Furthermore, study results were considered *negative* if one of three conditions were met. First, if a task-contingency was changed and participants' behavior was now always in line with this novel contingency. Second, if a self-generated or socially-provided rule was altered, and participants' behavior was now always in accordance with this novel rule. Third, if both a task-contingency *and* rule was changed, and participants' behavior was now always in line with this novel contingency *and* rule.

To cast votes for the second research question ("*Do adults suffering from psychological problems display a larger RBIE compared to their non-clinical counterparts?*"), we first assessed whether there was evidence supporting the RBIE. This was achieved in the same way as outlined above. If evidence for the effect was found, we subsequently examined if it was larger (in absolute terms) in the (sub-)clinical groups, compared to their non-clinical counterparts. If this was the case, then the study results would be categorized as *positive*. If these results were in the opposite direction, we would categorize them as *negative*.

Assessment of Risks of Bias

We, additionally, scrutinized the internal validity of the included studies. This examination involved assessing risks of bias using the Cochrane Collaboration tool for assessing risks of bias (Higgins & Altman, 2008) and the Office of Health Assessment and Translation (OHAT) Risk of Bias Rating Tool (NTP, 2015). Risks of bias can be defined as those aspects of a study design that can distort the conclusions that can be drawn from it. For the present review, we evaluated five potential risks of bias: *selection, exclusion, performance, detection, and reporting bias*. Note that these biases do not cover all risks of bias that are described in the Cochrane Collaboration and OHAT risks of bias tools. Indeed, given that these tools were not originally developed for assessing risks of bias in experimental-behavioral research, we selected and reformulated those risks of bias that we deemed relevant for evaluating such work.

For each of the studies, judgments of risks of bias (coded in terms of 'high', 'low', or 'unclear' risk of bias) were made in the following ways. To examine the possibility that there were systematic differences between the baseline characteristics of the groups that were compared (i.e., a *selection bias*), we examined: 1) the adequateness of a study's sequence generation procedure, 2) whether the experimental group to which participants were allocated to was concealed, 3) participants' past experiences with the experiment, and 4) the possibility that

they were misclassified to experimental groups. Furthermore, to assess the likelihood of an *exclusion bias* (i.e., systematic differences in the exclusion of participants from a study) we evaluated the possibility that there were systematic differences between groups with regard to the amount, nature, and handling of missing outcome data. To determine the risk of a *detection bias* (i.e., systematic differences between groups in how outcomes are determined) we evaluated: 1) the validity and reliability of the outcome assessment methods, 2) the adequateness of the outcome assessments, 3) the adequateness of the methods that were used to determine sample sizes and 4) the adequateness of the methods used to analyze the results. Judgments concerning *performance biases* (i.e., systematic differences between groups in how they were treated or exposed to factors other than the manipulation of interest) were made by examining whether: 1) the experimental contexts were standardized, 2) participants were informed about the study objectives, and 3) researchers and/or participants were informed about the experimental group to which participants were allocated to. Finally, to assess the possibility of a *reporting bias* (i.e., systematic differences between reported and unreported findings) we assessed potential discrepancies between the outcomes that were specified prior to the study and those that were eventually reported.

Assessment of External Validity

To determine the external validity of each of the included studies, we examined whether a study adequately described its eligibility criteria (in terms of age, sex, and diagnosis), the demographics of its sample, its study setting, its recruitment procedure, and the experimental manipulations that it used per experimental group.

Results

Summaries of Included Studies

For more information about the included studies, see Appendix S1 which contains summaries of all the included studies. These summaries are structured according to those studies that were deemed eligible to address the first (k = 20) and second research question (k = 1). There are two points worth noting about these summaries. First, they only include descriptions of those results that were relevant for the current research questions. As such, these summaries may contain less information than provided in the original study reports. Second, whenever it is mentioned that there is a difference between groups, this denotes an absolute and not a statistically significant difference.

Qualitative Synthesis: Source, Study, Task and Sample Characteristics

Source characteristics. The majority of the studies were written by a first author who did not work in the USA at the time of publication (i.e., Belgium [k = 3], Canada [k = 4], France [k = 2], Norway [k = 2], Switzerland [k = 1], USA [k = 9]) and most studies were published in the 2000s (k = 12).

Study characteristics. In the majority of the included studies, participants completed a conditional discrimination task (k = 14). In all of the studies, participants were allocated to one of the experimental groups, and conclusions about the RBIE were drawn by comparing the performances between these groups after a contingency change. Most of these studies examined the RBIE by examining how rules affected adaptation to changes (k = 11) or reversals (k = 6) in the non-instructed task-contingencies. See Table 1 for an overview of the study characteristics for each included study.

----------INSERT TABLE 1 HERE----------

Task characteristics. In each of the 21 included studies, a description was provided of the precise instructions or rules that were used. Seventeen of these studies reported how they manipulated their rules or instructions. In 16 of these cases, this was via written text (five of these studies also provided additional oral rules or instructions). The majority of the studies used socially-generated rules (k = 19; five of these studies also used self-generated rules), intermittent reinforcement schedules (k = 15; two of these studies also combined such schedules with continuous reinforcement schedules) and tasks that required simple discrete responses (k = 14; in two of these studies discrete choice responses were also required). In 18 out of the 21 studies, points were used as consequential stimuli which were often exchangeable for a monetary reward (k = 10 out of 18). Of those studies that reported whether a contingency change was announced (k = 9), seven of them stated that this was not the case (i.e., it was unannounced). Only one of the studies provided a description of the experimenter. Seven studies provided information about the presence of the experimenter. Of those studies, five stated that s/he was not present during the experiment. See Tables 2 and 3 for an overview of the task characteristics for each included study.

----------INSERT TABLE 2 HERE----------

----------INSERT TABLE 3 HERE----------

Sample characteristics. On average, 58 participants were included in the analyses (SD = 33 and range: 21-150). The mean age of participants was 20 (SD = .16) and the average number of females was 34 (SD = 25). Note, however, that these values were based on the two and six studies that reported the mean age and gender proportions of the samples that were included for analyses, respectively. Twenty out of the 21 studies used convenience samples, whereas only one study used students that were selected based on the presence or absence of sub-depressive symptomatology (i.e., Baruch et al., 2007)[3].

[3] Note that we did not use the schizophrenic patients group from the Monestès et al. (2014) study to address our second research question because it had fewer than ten participants within each experimental group.

Quantitative Synthesis: Vote Counting

To address Research Question 1 ("*Is there evidence for the rule-based insensitivity effect in adults?*") votes were only cast for the 11 out of the 20 studies that included a no-instructions group as a comparison group. These votes indicated that the results of each of these 11 studies were positive. No judgments could, however, be made for the one study that was relevant for addressing Research Question 2 ("*Do adults suffering from psychological problems demonstrate larger levels of the RBIE compared to their non-clinical counterparts?*"), because this study did not include a no-instructions group. For an overview of the vote-counting results for both research questions see Table 4.

----------INSERT TABLE 4 HERE----------

Assessments of Risks of Bias

Most of the included studies did not report the necessary information to assess all relevant domains of risks of selection, performance, exclusion, and detection bias. Nevertheless, the following can be said about those study aspects that we could draw conclusions about. Of the eleven out of the 21 studies that used a no-instructions group as a comparison group, none assessed the possibility that this group followed similar rules as the rules groups during the experiment. As a result, it could be that in these studies participants were misclassified to experimental groups. That is, there remains a possibility that participants were inaccurately thought to belong to a comparison group while in fact their behavior was actually governed by rules similar to those manipulated in the experimental groups. Furthermore, for the remaining domains, we argued that there were low risks of bias. Indeed, we argued that there was a low risk of reporting bias, seeing as there was a correspondence between the outcomes that were specified prior to the study and those that were actually reported. With respect to standardization of the experimental contexts, we argued that there was a low probability that the experimental groups were treated differently (performance bias). We also argued that there was a low probability that the methods that were used to assess the study outcomes were invalid or unreliable, and that the experimental groups differed with respect to how these outcomes were assessed (detection bias). See Appendices S2 and S3 for an overview of the judgments that were made for each aspect or domain of a study that could lead to a risk of bias.

Assessment of External Validity

The majority of those included studies that were relevant for examining our first research question ("*Is there evidence for the rule-based insensitivity effect in adults?"*) (k = 20) did not report all relevant demographics (i.e., mean age, sex, and education level) of their samples (k = 13) nor their recruitment procedure (k = 13). Most of these studies (k = 16), however, explicitly described the setting in which the experiment took place, and all of them provided a detailed description of the experimental manipulations per group. The one study that was relevant for examining our second research question (*"Do adults suffering from psychological problems*

display a larger RBIE compared to their non-clinical counterparts?"), selected participants based on the presence or absence of sub-clinical symptoms of depression, reported the eligibility criteria that they used, the demographics of their sample, and the experimental manipulations per group. Nevertheless, this study did not provide information about the experimental setting nor the procedure used to recruit participants.

Discussion

Rule-following is an essential human ability which can allow people to contact certain consequences more quickly and efficiently. Yet it has been argued that, under some conditions, this ability can also undermine people's sensitivity to other environmental contingencies (i.e., RBIE) and can lead to a wide range of clinical problems. Despite the presumed importance of this effect for our understanding of human behavior in general and human suffering in particular, to date, no systematic review has been carried out of the experimental work that has examined these claims. To this end, the present study systematically reviewed the RBIE literature to determine: 1) if there is evidence for the RBIE in adults and 2) if this effect is larger in adults suffering from psychological problems compared to their non-suffering counterparts. In addition, we investigated how 3) different operationalizations of the RBIE, and 4) the external validity and risks of bias of the experimental work investigating this effect, might influence the conclusions that can be drawn from the current systematic review.

Our results can be summarized as follows: (1) there is preliminary evidence for the idea that adults demonstrate the RBIE; (2) at present, there is no evidence to support the claim that psychological problems moderate the RBIE in adults; (3) similar procedures and tasks have been used to examine the RBIE, however, their precise parameters differed across studies; and (4) most studies did not report sufficient information to evaluate all relevant aspects concerning their external validity and risks of bias. In the following sections, we will elaborate on each of the above-described points and their implications for our understanding of this effect.

Evidence for the RBIE

Remarkably, only 11 out of the 20 studies that were deemed relevant for addressing our first research question ("*Is there evidence for the rule-based insensitivity effect in adults?"*) were eligible for vote-counting, because they included a no-instructions group (as a comparison group). Of these studies, the results showed that after the contingency change, the rule groups were more inclined to demonstrate behavior that was reinforced before the change, compared to their non-instructed counterparts. At first glance, this seems to suggest that when adults are asked to follow initially accurate rules, they experience more difficulties adapting to changes in contingencies (compared to when they are not asked to follow such rules). Nevertheless, the risk of bias assessments showed that such a conclusion may be premature because none of the 11 included studies assessed whether their no-instructions groups functioned as adequate comparison groups. That is, none of these studies examined if, during the experiment,

participants in their comparison group did not follow rules about the task-contingencies that were similar to those followed by the rule groups. As a result, this made it difficult to determine whether the effects that were observed in the rules groups could be attributed to the rules or instructions that were manipulated in those experiments.

Despite the fact that we found preliminary evidence for the RBIE in all 11 studies, it is important to acknowledge that there might be variables that increase or decrease the likelihood of observing this effect. For instance, according to past work, the RBIE might be less likely to occur if the experimenter is not physically present (e.g., Kroger-Costa & Abreu-Rodrigues, 2012), participants are provided with inaccurate as opposed to accurate instructions before a contingency change occurs (e.g., Hojo, 2002), and if the consequences for behaving in line with the actual task-contingencies outweigh those of following the rule (Donadeli & Strapasson, 2015). Unfortunately, a systematic examination of potential moderators of the RBIE (besides the moderating impact of the absence/presence of psychological problems) was beyond the scope of this systematic review. Nonetheless, we deem such an examination vital as it might further our understanding of the robustness of this effect. As such, we recommend that future work systematically examines those variables that might decrease or increase the RBIE.

Psychological Problems and the RBIE

Despite the key role that the RBIE has been argued to play in psychological problems, only one of the included studies was deemed relevant for examining this idea. However, given that this study did not include a no-instructions group, no judgments could be made about the extent to which evidence was found for the RBIE, and whether psychological problems moderated this effect. This suggests that there is currently no evidence available to draw firm conclusions about the relationship between psychological problems and the RBIE in adults. Furthermore, even if we evaluated the peer-reviewed journal articles (n = 69) which examined the RBIE but were omitted because they: (a) used samples smaller than 10, (b) samples from non-adult populations, (c) used non-experimental designs, and/or (d) did not manipulate rules or include a contingency change, we still failed to identify many relevant studies. Indeed, such a revised search only resulted in an additional four studies: two studies that investigated the impact of sub-clinical depressive symptoms in adolescents (McAuliffe et al., 2014 [Experiments 1 and 2]), one study that examined that of ADHD in children (Kollins, Lane, & Shapiro, 1997) and another study that examined that of schizophrenia in samples smaller than 10 (Monestès et al., 2014). We, therefore, strongly recommend that more work is conducted on the relationship between the RBIE and psychological problems to better inform clinical theory and treatment.

When carrying out such work, researchers should also explore certain variables that could moderate this effect in clinical groups. For instance, it might be that clinical groups (e.g., arachnophobic) are more insensitive to contingency changes if they follow pathology-relevant (e.g., "*If you want to remain alive, always avoid places where there could be spiders*") but not pathology-irrelevant rules (e.g., "*to gain points press the blue button*"). Likewise, it is possible

that different clinical groups (people suffering from psychosis vs. depression) demonstrate different levels of the RBIE because of differences in the origins (generated by imaginary agents vs. self-generated) of the rules they follow. Another possibility is that variations in the elements of the rules (i.e., the described stimuli [all spiders vs. tarantulas], responses [avoiding spiders vs. attacking them], and contexts [all spider habitats vs. the basement]), might contribute to differences in how people suffering from similar conditions (e.g., arachnophobia) adapt to contingency changes. We believe that such an endeavor would be useful because it could aid clinicians in developing more targeted treatments.

Operationalization of the RBIE

Our coding of task and study characteristics revealed that although most of the included studies used similar tasks and procedures, the precise parameters that were involved often differed. Specifically, many studies used conditional discrimination tasks during which participants could initially gain points if they followed the rules they received from the experimenter. In most of these studies, the task-contingencies were subsequently altered after a number of trials so that the previously effective rules were rendered ineffective. To illustrate, consider Kissi et al.'s (2018) Matching-To-Sample (MTS) task. This task consisted of two experimental phases. On every trial, participants were presented with four images. One image – called the 'sample stimulus' – was presented at the top of the screen and always consisted of three identical symbols or letters (e.g., TTT). Three other images – called the comparison stimuli – were presented at the bottom of the screen. One of these images had two symbols or letters that were identical to the sample stimulus (e.g., TT%; most-like comparison stimulus), another had one symbol or letter identical to the sample stimulus (e.g., T%%; moderate-like comparison stimulus), while the third had no symbols or letters in common with the sample stimulus (e.g., %%%; least-like comparison stimulus). During the first phase of the experiment, participants could obtain points if they selected the comparison stimulus that was most-like the sample stimulus. However, during the second phase of the experiment, the task-contingencies were changed. Now, participants gained points whenever they selected the comparison stimulus that was least-like the sample stimulus. To examine the RBIE, some participants were given instructions telling them how to gain points in the task, whereas others had to learn about the task-contingencies via trial-and-error. This task is a conditional discrimination task because reinforcement for responses was conditional upon the characteristics of the sample stimulus.

Critically, despite the fact that most included studies used similar tasks, the precise stimuli (tones vs. images) that were used, the point in time in which the contingency change occurred (e.g., after two vs. three blocks), and the study outcomes (e.g., latencies vs. rate or accuracy of responding) often differed between studies. Generally speaking, if reliable evidence is found for a phenomenon, such variations are often viewed as a potential advantage because they enhance the generalizability of a study's findings. Yet given that, in our opinion, it is unclear whether the RBIE was adequately assessed in any of the included studies in this review,

we believe that this idea cannot be applied to our findings (see the previous sections “Evidence for the RBIE” and “Psychological Problems and the RBIE”).

External Validity and Risks of Bias

The results revealed that many studies did not report all relevant demographics of their samples, how they were recruited, if the contingency changes were announced, and if the experimenter was present during the experiment. In addition, no study provided sufficient information to assess all domains of potential risks of bias. Taken together, this suggests that the reports of the included studies did not provide sufficient information to evaluate all coding items assessing their external and internal validity. The lack of such information is particularly problematic in the context of systematic reviews because it limits the conclusions that can be drawn from it. As such, we strongly recommend that, in future work, researchers report all information about their study that may enable readers to more readily draw conclusions about its external and internal validity (see Schulz, Altman, & Moher, 2010 for guidelines).

Other Considerations

In many of the studies, there was the implicit assumption that when people were asked to follow accurate rules, their behavior would be exclusively governed by those rules, and that if this was not the case, their actions would be exclusively guided by the task-contingencies. We would argue that such a reasoning might be problematic for two reasons (for similar arguments see Hayes, Brownstein, Haas, & Greenway, 1986). First, previous work suggests that when humans are *not* provided with rules they rarely demonstrate purely contingency-shaped behavior. Instead, they often generate and use their own rules about how they should behave in a particular context, based on their (trial-and-error) experiences in that context (Rosenfarb, Newland, Brannon & Howey, 1992; Shimoff, Matthews & Catania, 1986). Second, such an interplay between environmental contingencies and rules may have also impacted the behavior of the rule groups that were described in the reviewed studies. Indeed, a closer look at the results of these studies showed that when behavior was considered rule-governed, it was rarely ever the case that participants consistently stuck to the rules they were told to follow. Rather, the results suggest that participants sometimes engaged with the task in ways that were not specified by these rules. There could be two possible explanations for this finding. A first possibility is that these deviations from the rules were unintentional and as such reflected erroneous responding. A second possibility is that instances in which participants discarded the rules that they were told to follow, actually constituted intentional attempts to explore instead of exploit the task-contingencies (Berger-Tal, Nathan, Meron, & Saltz, 2014).

If the latter possibility is valid as well as the possibility that rules governed the behavior of the no-instructions groups, then this might suggest that comparisons between instructed and non-instructed groups might not inform us about the effects of rule-governed vs. contingency-shaped behavior *per se*. Indeed, such comparisons might then rather inform us about the *relative*

degree to which socially-provided rules vs. environmental contingencies and self-generated rules vs. environmental contingencies influenced the behavior of the instructed and non-instructed groups, respectively. Yet given that we could not assess the plausibility of this assertion in the current study, this idea remains speculative. We, therefore, recommend that future work examines its validity so that we can gain a better understanding of how the RBIE should be conceptualized (e.g., as an insensitivity of behavior to other contingencies due to a *stronger* reliance on socially-generated rules than environmental contingencies).

Finally, to the best of our knowledge, there is currently no consensus about how contingency insensitive and sensitive behavior should be measured. Indeed, if anything, the implicit assumption is that behavior is contingency insensitive if it is not in line with a contingency, whereas it is contingency sensitive if it corresponds with a contingency. We believe that although such operational definitions can be useful in some respects, they lack the precision that is needed to measure these behaviors in a uniform and unambiguous manner. Indeed, given the broad and descriptive nature of these definitions, much variation can exist between studies in how they measure contingency sensitive and insensitive behavior. We believe that, although this is not an issue per se, it can become problematic when one wants to draw general conclusions across studies. We, therefore, recommend that future work offers more precise operational definitions of contingency sensitive and insensitive behavior.

Limitations

Several factors should be taken into account when interpreting our results. First, to determine whether or not behavior was in line with a previously effective rule and/or a novel contingency we used a liberal criterion. That is, we considered participants' behavior to be in line: 1) with a previously effective rule if they demonstrated behavior that corresponded with this rule on at least a few trials, and 2) with a novel contingency and/or rule if they always behaved in line with this contingency and/or rule. As a consequence, it possible that if a different criterion were used, other findings would have emerged. Second, we opted for vote-counting for our quantitative research synthesis, which unlike the standard meta-analytic approach does not provide information about the magnitude of the observed effects (Koricheva & Gurevitch, 2013). Nevertheless, to gain some insight into these effects, we conducted a random effects model meta-analysis using those studies that reported sufficient statistical information. This analysis was based on six studies including a total of 377 participants (i.e., Haas & Hayes, 2006; Harte et al., 2017 [Experiment 2], Kissi et al., 2018; Kudadjie-Gyamfi & Rachlin, 2002; Monestès et al., 2017; Monestès et al., 2014). It revealed a significant effect size of .76 (Cohen's *d* for independent samples; 95% CI [.41 – 1.12]; $p < .001$) indicating that participants had *far more* difficulties adapting to a contingency change if, prior to the change, they received a rule as opposed to no rule. Third, across *all* studies that were deemed eligible for vote-counting, preliminary evidence was found for the RBIE. This was surprising, given that, in general, the likelihood of observing the same effect across all studies in a systematic review is rather low (Thornton & Lee, 2000). Usually, when such an overrepresentation of positive effects is

observed, it is assumed that this might be due to publication bias, i.e., journals' preference for publishing positive over negative findings (Joober, Schmitz, Annable & Boksa, 2012; Thornton & Lee, 2000). Publication bias is particularly problematic in the context of systematic reviews, because it can lead to an overestimation of the existence of a particular effect. Therefore, we recommend the reader to take this bias into account when interpreting the findings of our systematic review. Finally, we adopted pre-defined inclusion and exclusion criteria which inevitably limited the scope of the review and as such the potential conclusions that can be drawn from it. For instance, we only considered peer-reviewed journal articles that examined one instance of the RBIE and one potential moderator of this effect in adult populations. Similarly, we only included experiments with groups that contained at least 10 participants, which led us to discard naturalistic studies and studies that adopted a single-subject methodology.

Conclusions

For several decades now, the RBIE has been argued to play an important role in human behavior in general and psychological suffering in particular. Yet despite its widespread appeal, the results of this systematic review suggest that strong claims about its existence and role in psychological suffering are currently unsupported and thus far unwarranted. Indeed, at present, only preliminary evidence exists concerning the RBIE in adults and no strong evidence is available to draw conclusions about its role in the development and maintaince of psychological suffering in adults. We, therefore, recommend that more systematic research is conducted on the RBIE so that future work can better evaluate the relevance of this effect for our understanding of human behavior and psychological suffering.

References

Baruch, D. E., Kanter, J. W., Busch, A. M., Richardson, J. V, & Barnes-holmes, D. (2007). The differential effect of instructions on dysphoric and nondysphoric persons. *The Psychological Record, 57(4)*, 543–554.

Blackledge, J. T., & Drake, C. E. (2013). Acceptance and commitment therapy: Empirical and theoretical considerations. In S. Dymond & B. Roche (Eds.), *Advances in relational frame theory and contextual behavioral science: Research and application* (pp. 219–252). Oakland, CA: New Harbinger.

Berger-Tal, O., Nathan, J., Meron, E., Saltz, D. (2014). The Exploration-Exploitation Dilemma: A Multidisciplinary Framework. PLOS ONE, 9(4), e95693. Doi: 10.1371/journal.pone.0095693

Bushman, B. J., & Wang, M. C. (2009). Vote-Counting Procedures in Meta-Analysis. In H. Cooper, L. V. Hedges, & J. F. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed., pp. 207-220). New York: Russel Sage Foundation.

Cerutti, D. (1994). Compliance with instructions: Effects of randomness in scheduling and monitoring. *The Psychological Record, 44*, 1–6.

Cerutti, D. T. (1991). Discriminative versus reinforcing properties of schedules as determinants of schedule insensitivity in humans. *The Psychological Record*, 41, 51-67.

Deeks, J. J., Higgins, J. P. T., Altman, D. G. (Eds.) (2008). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins J. P. T. & Green S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.0, pp. 243-296).Chichester, UK: John Wiley & Sons.

Dixon, M. R., Hayes, L. J., & Aban, I. B. (2000). Examining the roles of rule following, reinforcement, and preexperimental histories on risk-taking behavior. *Psychological Record, 50*(4), 687-704. Doi:10.1007/BF03395378

Donadeli, J. M., & Strapasson, B. A. (2015). Effects of Monitoring and Social Reprimands on Instruction-Following in Undergraduate Students. *Psychological Record, 65*(1), 177-188. Doi:10.1007/s40732-014-0099-7

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics, 19*(4), 27–29. Doi:10.1590/2176-9451.19.4.027-029.ebo

Haas, J. R., & Hayes, S. C. (2006). When Knowing You Are Doing Well Hinders Performance : Exploring the Interaction Between Rules and Feedback. *Journal of Organizational Behavior Management, 26*(1-2), 91-111. Doi:10.1300/J075v26n01_04

Harte, C., Barnes-Holmes, Y., Barnes-Holmes, D., & McEnteggart, C. (2017). Persistent Rule-Following in the Face of Reversed Reinforcement Contingencies: The Differential Impact of Direct Versus Derived Rules. *Behavior Modification, 41*(6), 743-763. Doi:10.1177/0145445517715871

Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, multiple schedules, and extinction: Distinguishing rule-governed from schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 46*(2), 137–147.

Doi:10.1901/jeab.1986.46-137
Hayes, S. C., & Gifford, E. V. (1997). The trouble with language: Experiential Avoidance, Rules, and the Nature of Verbal Events. *Psychological Science, 8*(3),170-173. Doi:10.1111/j.1467-9280.1997.tb00405.x
Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). The Annual Review of Clinical Psychology is online. *Annu. Rev. Clin. Psychol, 7*, 141–168. doi :10.1146/annurev-clinpsy-032210-104449
Higgins J. P. T. & Altman D. G. (Eds.) (2008). Chapter 8: Assessing risk of bias in included studies. In: Higgins J. P. T., & Green S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.0, pp. 187-241). Chichester, UK: John Wiley & Sons.
Hojo, R. (2002). Effects of instructional accuracy on a conditional discrimination task. *The Psychological Record, 52*(4), 493–506.
Hughes, S., & Barnes-Holmes, D. (2015). Relational Frame Theory: The Basic Account. In R. Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioural science* (pp. 178–226). West Sussex, UK: Wiley-Blackwell.
Joober, R., Schmitz, N., Lawrence, A., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *J Psychiatry Neurosc, 37*(3), 149-152.
Joyce, J. H., & Chase, P. N. (1990). Effects of response variability on the sensitivity of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 54*, 251-262.
Kissi, A., Hughes, S., De Schryver, M., De Houwer, J., & Crombez, G. (2018). Examining the Moderating Impact of Plys and Tracks on the Insensitivity Effect: a Preliminary Investigation. *The Psychological Record, 68*(4), 431-440. Doi:10.1007/s40732-018-0286-z
Kissi, A., Hughes, S., Mertens, G., Barnes-Holmes, D., De Houwer, J., & Crombez, G. (2017). A Systematic Review of Pliance, Tracking, and Augmenting. *Behavior Modification, 41*(5), 683-707. Doi:10.1177/0145445517693811
Kollins, S. H., Lane, S. D., & Shapiro, S. K. (1997). Experimental analysis of childhood psychopathology: A laboratory matching analysis of the behavior of children diagnosed with Attention-Deficit Hyperactivity Disorder (ADHD). *Psychological Record, 47*, 25-44.
Koricheva, J. & Gurevitch, J. (2013). Place of Meta-analysis among Other Methods of Research Synthesis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.). *Handbook of Meta-Analysis in Ecology and Evolution*. (pp. 3–13). Princeton, NJ: Princeton University Press.
Kroger-Costa, A.,& Abreu-Rodrigues, J. (2012). Effects of historical and social variables on instruction following. *The Psychological Record,*62(4), 691–705.
Kudadjie-Gyamfi, E., & Rachlin, H. (2002). Rule-governed versus contingency-governed behavior in a self-control task: Effects of changes in contingencies. *Behavioural Processes, 57*(1), 29-35. Doi:10.1016/S0376-6357(01)00205-4
Lefrancois, J. R., Chase, P. N., & Joyce, J. H. (1988). The effects of a variety of instructions on human fixed-interval performance, *Journal of the Experimental Analysis of Behavior, 49*(3), 383–393.
McAuliffe, D., Hughes, S., & Barnes-Holmes, D. (2014). The dark-side of rule governed

behavior: An experimental analysis of problematic rule-following in an adolescent population with depressive symptomatology. *Behavior Modification*, *38*, 587–613. Doi:10.1177/0145445514521630

McCracken, L. M. (2005). *Contextual cognitive-behavioral therapy for chronic pain*. Seattle, WA: International Association for the Study of Pain.

McCracken, L. M., Carson, J. W., Eccleston, C., & Keefe, F. J. (2004). Acceptance and change in the context of chronic pain. *Pain, 107*, 159-166. Doi:10.1016/j.pain.2004.02.006

Miller, J. R., Hirst, J. M., Kaplan, B. A., DiGennaro Reed, F. D., & Reed, D. D. (2014). Effects of Mands on Instructional Control: A Laboratory Simulation. *The Analysis of Verbal Behavior, 30*(2), 100-112. Doi:10.1007/s40616-014-0015-x

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. The PRISMA Statement. *PLoS Med 6(7)*: e1000097. Doi:10.1371/journal.pmed1000097

Monestès, J. L., Greville, W. J., & Hooper, N. (2017). Derived insensitivity: Rule-based insensitivity to contingencies propagates through equivalence. *Learning and Motivation*, *59*, 55–63. Doi:10.1016/J.LMOT.2017.08.003

Monestès, J. L., Villatte, M., Stewart, I., & Loas, G. (2014). Rule-Based Insensitivity and Delusion Maintenance in Schizophrenia. *The Psychological Record*, *64*(2), 329–338. Doi:10.1007/s40732-014-0029-8

Ninness, H. A., C. & Ninness, S. K. (1998). Superstitious math performance: Interactions between rules and scheduled contingencies. *The Psychological Record*, *48*(1), 45–62.

NTP (2015). *Handbook for conducting a literature-based health assessment using OHAT approach for systemic review and evidence integration*. Retrieved from https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf

Otto, T. L., Torgrud, L. J., & Holborn, S. W. (1999). An operant blocking interpretation of instrcuted insensitivity to schedule contingencies. *Psychological Record, 49*, 663-684. Doi:10.1007/BF03395334

Rosenfarb, I. W., Newland, M. C., Brannon, S. E., Howey, D. S. (1992). Effects of self-generated rules on the development of schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 58*(1), 107-121. Doi: 10.1901/jeab.1992.58-107

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*. Doi:10.1186/1741-7015-8-18

Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior*, *36*(2), 207–220. Doi:10.1901/jeab.1981.36-207

Souza, A. S., Pontes, T. N. R., & Abreu-Rodrigues, J. (2012). Varied but not necessarily random: Human performance under variability contingencies is affected by instructions. *Learning and Behavior*, *40*(4), 367–379. Doi:10.3758/s13420-011-0058-y

Svartdal, F. (1989). Shaping of rule-governed behaviour. *Scandinavian Journal of Psychology,*

30(4), 304-314. Doi:10.1111/j.1467-9450.1989.tb01093.x

Svartdal, F. (1995). When feedback contingencies and rules compete: Testing a boundary condition for verbal control of instrumental performance. *Learning and Motivation*, 221–238.

Thornton, A., & Lee, P. (2002). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology, 53*(2), 207-216.

Torgrud, L. J., Holborn, S. W., & Zak, R. D. (2006). Determinants of human fixed-interval performance following varied exposure to reinforcement schedules. *Psychological Record, 56*(1), 105-133. Doi:10.1007/BF03395540

Törneke, N., Luciano, C., & Salas, S. V. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology and Psychological Therapy, 8*(2), 141-156.

Van Ryckeghem, D. M. L., Van Damme, S., Eccleston, C., & Crombez, G. (2018). The efficacy of attentional distraction and sensory monitoring in chronic pain patients: A meta-analysis. *Clinical Psychology Review, 59*, 16-29.

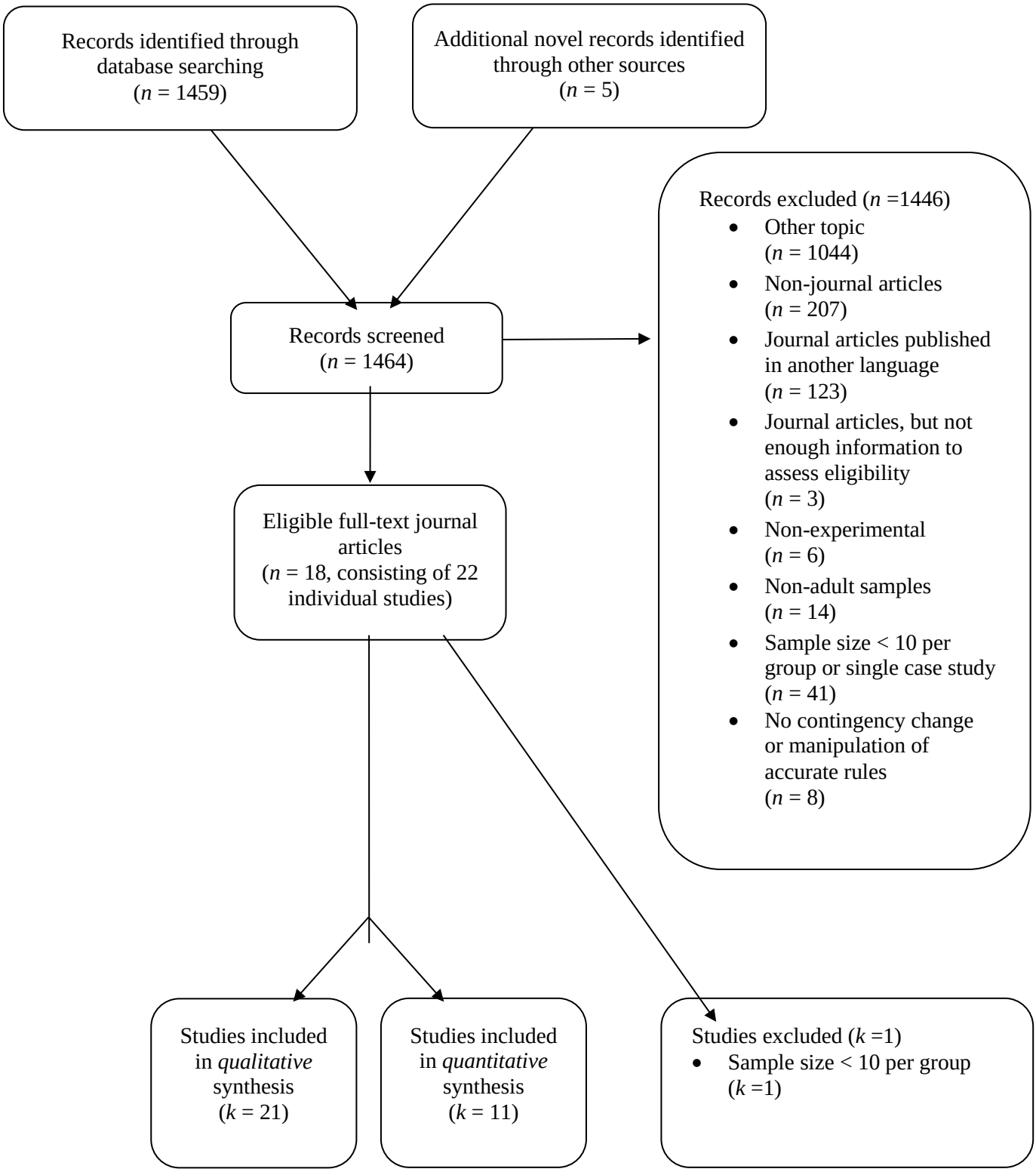

Figure 1. Flow diagram of the study selection process.

.

Table 1:
Coded study characteristics

	Type of task	Experimental design	Procedure	Analytic method
Baruch et al. (2007)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Cerutti (1991)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Cerutti (1994)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Dixon et al. (2000)	Gambling task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Haas and Hayes (2006)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change

Harte et al. (2017 – Experiment 1)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Harte et al. (2017 – Experiment 2)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Hayes et al. (1986)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Kissi et al. (2018)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Kudadjie-Gyamfi and Rachlin (2002)	Distributed choice paradigm where reinforcement could be increased if participants minimized the delay between a choice and its outcome	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
LeFrancois et al. (1988)	Task in which reinforcement was dependent upon button presses	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change

730

Monestès et al. (2017)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Monestès et al. (2014)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies reversal	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Otto et al. (1999 – Experiment 1)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Otto et al. (1999 – Experiment 2)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Shimoff et al. (1981)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Souza et al. (2012)	Task in which participants had to generate three-digit sequences that met a variability criterion in order to receive reinforcement	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Svartdal (1989)	Task in which participants had to count clicks and insert the number of clicks that they thought they heard in order to receive reinforcement	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Svartdal (1995 – Experiment 2)	Conditional discrimination task	Participants were allocated to one of the experimental groups	Non-instructed and instructed contingency change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change
Torgrud et al. (2006 –	Task in which reinforcement was dependent upon	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing

	Experiment 1)	participants' pattern of key presses			performances between groups after a contingency change
	Torgrud et al. (2006 – Experiment 2)	Task in which reinforcement was dependent upon participants' pattern of key presses	Participants were allocated to one of the experimental groups	Non-instructed task contingencies change	Conclusions about RBIE are drawn by comparing performances between groups after a contingency change

731
732
733
734

Table 2:
Coded study characteristics

	Report of exact rules/instructions used	Rule-delivery	Rule-generation	Reinforcement schedule(s)	Behavioral responses
Baruch et al. (2007)	Yes	Written	Socially-generated	Continuous	Discrete choice responses
Cerutti (1991)	Yes	Written	Self-generated	Intermittent	Discrete simple and discrete choice responses
Cerutti (1994)	Yes	Written	Self-generated	Intermittent	Discrete simple and discrete choice responses
Dixon et al. (2000)	Yes	Written	Socially-generated	Intermittent	Discrete simple responses
Haas and Hayes (2006)	Yes	Written and orally	Socially –and self-generated	Continuous and intermittent	Discrete simple responses
Harte et al. (2017 – Experiment 1)	Yes	Unclear	Socially –and self-generated	Continuous	Discrete choice responses

Harte et al. (2017 – Experiment 2)	Yes	Unclear	Socially –and self-generated	Continuous	Discrete choice responses
Hayes et al. (1986)	Yes	Written and orally	Socially-generated	Intermittent	Discrete simple responses
Kissi et al. (2018)	Yes	Written	Socially-generated	Continuous	Discrete choice responses
Kudadjie-Gyamfi and Rachlin (2002)	Yes	Written	Socially –and self-generated	Continuous and conditional	Discrete choice responses
LeFrancois et al. (1988)	Yes	Written	Socially-generated	Intermittent	Discrete simple responses
Monestès et al. (2017)	Yes	Written and orally	Socially-generated	Intermittent	Discrete choice responses
Monestès et al. (2014)	Yes	Orally	Socially –and self-generated	Intermittent	Discrete simple responses
Otto et al. (1999 – Experiment 1)	Yes	Written and orally	Socially-generated	Intermittent	Discrete simple responses

739

Otto et al. (1999 – Experiment 2)	Yes	Written and orally	Socially-generated	Intermittent	Discrete simple responses
Shimoff et al. (1981)	Yes	Written	Socially-generated	Intermittent	Discrete simple responses
Souza et al. (2012)	Yes	Written	Socially-generated	Continuous	Complex response (i.e., three-digit combinations)
Svartdal (1989)	Yes	Unclear	Socially-generated	Continuous and intermittent	Discrete simple responses
Svartdal (1995 – Experiment 2)	Yes	Unclear	Socially-generated	Intermittent	Discrete simple responses
Torgrud et al. (2006 – Experiment 1)	Yes	Both	Socially-generated	Intermittent	Discrete simple responses
Torgrud et al. (2006 – Experiment 2)	Yes	Both	Socially-generated	Intermittent	Discrete simple responses

Table 3:
Coded study characteristics

	Consequential stimuli	Announcement of contingency change(s)	Description of experimenter	Presence of experimenter
Baruch et al. (2007)	Points that were exchangeable for a monetary reward	Unclear	Yes	No
Cerutti (1991)	Points that were exchangeable for a monetary reward and a tone	Unclear	No	Yes
Cerutti (1994)	Points	Unclear	No	Unclear
Dixon et al. (2000)	Chips that were exchangeable for extra credit points	Unannounced	No	No
Haas and Hayes (2006)	Points that were exchangeable for a monetary reward	Unannounced	No	Unclear
Harte et al. (2017 - Experiment 1)	Points	Unannounced	No	Unclear
Harte et al. (2017 - Experiment 2)	Points	Unannounced	No	Unclear
Hayes et al. (1986)	Points that were exchangeable for a monetary reward	Unclear	No	No

Kissi et al. (2018)	Points	Unannounced	No	No
Kudadjie-Gyamfi and Rachlin (2002)	Points that were exchangeable for a monetary reward and time delays	Unclear	No	Unclear
LeFrancois et al. (1988)	Points that were exchangeable for a monetary reward	Unclear	No	Unclear
Monestès et al. (2017)	Points	Unclear	No	Unclear
Monestès et al. (2014)	Points	Unannounced	No	Yes
Otto et al. (1999 - Experiment 1)	Points	Unclear	No	Unclear
Otto et al. (1999 - Experiment 2)	Points	Unclear	No	Unclear
Shimoff et al. (1981)	Points that were exchangeable for a monetary reward	Unclear	No	Unclear

746

Souza et al. (2012)	Points that were exchangeable for a monetary reward	Unannounced	No	Unclear
Svartdal (1989)	Unclear	Announced	No	Unclear
Svartdal (1995 - Experiment 2)	Sounds and lights	Announced	No	No
Torgrud et al. (2006 - Experiment 1)	Points that were exchangeable for a monetary reward	Unclear	No	Unclear
Torgrud et al. (2006 - Experiment 2)	Points that were exchangeable for a monetary reward	Unclear	No	Unclear

747

748 **Table 4:**

749 **Overview of vote-counting results**

Studies used to answer Research Question 1 ("Is there evidence for the rule-based insensitivity effect in adults")		
Type of change	**Experiment**	**Evidence for the RBIE**
Task-contingencies		
	Dixon et al. (2000)	+
	Haas & Hayes (2006)	+
	Harte et al. (2017 - Experiment 1)	Unclear
	Harte et al. (2017 - Experiment 2)	+
	Hayes et al. (1986)	+
	Kissi et al. (2018)	+
	Kudadjie-Gyamfi and Rachlin (2002)	+
	LeFrancois et al. (1988)	+
	Monestès et al. (2017)	+
	Monestès et al. (2014)	+
	Otto et al. (1999 - Experiment 2)	Unclear
	Shimoff et al. (1981 - Experiment 1)	+
	Souza et al. (2012)	+
	Svartdal (1989)	Unclear
	Torgrud et al. (2006 – Experiment 1)	Unclear
	Torgrud et al. (2006 – Experiment 2)	Unclear
Instructions		
	Cerutti (1991)	Unclear
	Cerutti (1994)	Unclear
	Otto et al. (1999 - Experiment 1)	Unclear
Task-contingencies and instructions		
	Svartdal (1995 - Experiment 2)	Unclear

Studies used to answer Research Question 2 ("Do adults suffering from psychological problems display a larger RBIE compared to their non-clinical counterparts?").		
Type of change	**Experiment**	**Evidence for a larger RBIE in the clinical group**
Task-contingencies		
	Baruch et al. (2007)	Unclear

Note. ‘+’ indicates that there was evidence for the RBIE. ‘-‘ indicates that participants in the rule-group(s) adapted to the change in the task-contingencies or instructions. ‘Unclear’ indicates that there was insufficient information to cast votes.

Appendix S1 : Summary of all the included studies

Studies included to answer Research Question 1. Cerutti (1991) examined the moderating effects of mixed-random, mixed-fixed, and fixed-time schedules on the manner in which participants adapted to changes in the reinforcement delivered for self-generated rules about the task-contingencies. Participants were presented with one of three schedules (a mixed-random time schedule [n = 10]; a mixed-fixed time schedule [n = 11]; a fixed-time (FT) 3.3 schedule [n= 10]). During each of these schedules they were asked to avoid the occurrence of tones by pressing one of two panels (A & B), and to earn points by guessing how they could prevent these tones (i.e., by generating rules about these contingencies). During the initial phase of the experiment, participants earned points if they indicated that they thought that pressing panel A rapidly prevented the tones. In the second phase, however, these contingencies changed so that now points were only earned for high-rate guesses for panel B. Note, that these points for guesses were not contingent upon the extent to which they accurately reflected the task-contingencies, but were randomly shaped. The results indicated that participants in the mixed schedules groups were inclined to demonstrate behavior that was in line with what they thought prevented the tones (e.g., pressing fast or slow), despite the non-corresponding contingencies, while this was not the case in the FT schedule group.

Once again in 1994, Cerutti investigated the effects of three different types of reinforcements schedules on participants' adaptation to changes in self-generated rules using a similar paradigm as in his 1991 study. The most essential procedural difference between both studies was that now participants were quasi-randomly assigned to either a random-interval (RI) 10 schedule (n = 20), fixed-interval (FI) 10 schedule (n = 20), or FI 10 schedule with videotaping (n = 20). The results suggested that when the reinforcement contingencies for the guesses (i.e., self-generated rules) were reversed (i.e., when high-rate guesses for panel B instead of panel A were reinforced), compliance with these reversed guesses was more likely under the FI schedule with videotaped performance and the RI schedule, compared to the FI schedule alone. The RI and FI schedule with videotaped performance, however, did not differ in the extent to which they adhered to the reversed guesses.

Dixon, Hayes, and Aban (2000) examined the effects of the accuracy of instructions on behavior, when the chances of receiving reinforcement were rendered low. Participants randomly received accurate (n = 15), inaccurate (n = 15) or no-instructions (n = 15) about how they should play a game of roulette. When these instructions were presented this was accompanied by payback percentages of p = .2, p = .8 or those that were fair. In the next phase of the experiment, the winning probabilities were all set to p = .2 and participants were given the opportunity to quit the game. Results showed that participants who received instructions were more likely to quit the game compared to those that were not given any instructions. This tendency was, furthermore, larger in the inaccurate compared to accurate instructions group, indicating that the former group behaved less in line with the new reinforcement schedule compared to the latter group. No

comparison could be made between the different winning probability groups, given that $N < 10$ within each of these groups.

Haas and Hayes (2006) examined the unique and combinatory effects of two types of verbal feedback: rule-following and task performance feedback, and the accuracy of rule-following feedback on participants' adaptation to changes in the task-contingencies. Participants were randomly allocated to one of six groups (10 in each group): the inaccurate rule-following feedback, accurate rule-following feedback, inaccurate rule-following + task performance feedback, accurate rule-following + task performance feedback, rule alone or minimal rule group. In each of these groups, participants had to move a shape on a screen through a grid to earn points. Before starting the task, all participants, except those in the minimal rule group, received accurate instructions about how they could earn points during Phase 1. Specifically, these participants were told that points could be earned by pressing the buttons slowly if the blue rectangle is lit and rapidly when the red rectangle is lit (both of which appeared on the screen below the grid). The reinforcement schedules that were in effect during Phase 1 were a Differential Reinforcement of Low rates (DRL) 6 schedule when the blue rectangle was lit and a Fixed-Ratio (FR) 18 schedule when the red rectangle was lit. Towards the end of Phase 1, participants received feedback about whether their behavior corresponded with the rules they received and/or their task performances (depending on their experimental group). This feedback was either accurate (in the accurate rule groups) or inaccurate (i.e., non-contingently positive in the inaccurate rule groups). During Phase 2, the task-contingencies changed so that now reinforcement was delivered according to a multiple FR 1 schedule when the blue rectangle was lit, and an FI yoked schedule (i.e., the interval reflected the average number of seconds that participants needed to respond 18 times during the last FR component) when the red rectangle appeared. The results indicated that, on average, participants failed to adapt to the changes in the reinforcement schedules fully. This was mainly the case in the accurate rule-following with task performance feedback group when the DRL 6 schedule changed to an FR 1 schedule, and the accurate rule-following feedback without task performance feedback group when the FR 18 schedule changed to an FI yoked schedule.

Across two experiments Harte, Barnes-Holmes, Barnes-Holmes, and McEnteggart (2017) examined how receiving a direct rule versus deriving a rule affected how participants adapted to changes in reinforcement contingencies. In Experiment 1, participants were randomly assigned to either a direct ($n = 25$) or a derived ($n = 44$) rule group. In Phase 1, participants completed a conditional discrimination task in which they initially always received points if they correctly matched stimuli according to their physical dissimilarities. In Phase 2, however, the task-contingencies were reversed so that now points could only be earned if participants correctly matched stimuli according to their physical similarities. The results showed that, of those participants that met the specific performance criteria, after the contingency reversal, both the direct and derived rule groups adhered to the rules that were effective prior to the reversal. This effect, however, appeared to be slightly larger in the direct compared to the derived rule group.

In Experiment 2, Harte and colleagues tried replicating this finding using a similar procedure as in Experiment 1, with two notable exceptions. First, participants now had more opportunities to follow the reinforced rule in Phase 1 than in Experiment 1 (10 trials in Exp. 1 vs. 100 in Exp. 2). Second, a comparison group was also included that did not receive rules about how to earn points, and as such had to detect the task-contingencies themselves. Twenty-five participants were assigned to this group, while the remaining participants were randomly allocated to the direct (n = 39) or derived (n = 76) rule groups. The results suggested that, of those participants that met the specific performance criteria, all groups were somehow inclined to demonstrate behavior that was reinforced before the contingency reversal. This tendency, however, appeared to be the largest in the direct rule group, followed by the derived rule group and then the comparison group.

Hayes et al. (1986) examined whether initially partially accurate (n = 13), accurate (n = 16) or no-instructions (n = 19) regarding appropriate rates of responding, influenced participants behavior during extinction. Irrespective of the instructions that were given, all participants could initially earn points if they pressed buttons according to a DRL 6 schedule when a yellow rectangle was lit, and FR 18 schedule when a blue square was lit. After a certain period, an extinction phase was introduced during which responses were no longer reinforced. The results showed that, on average, almost all participants continued to emit responses during extinction (i.e., after the task-contingency change). This was more so for the accurate instructions group compared to the partially accurate and no-instructions groups, and for the no-instructions group compared to the partially accurate instructions group.

Kissi et al. (2018) examined the moderating effects of two types of rules (plys and tracks [4]) on participants' adaptation to a task-contingency change. Participants were randomly assigned to one of three groups: a ply (n = 15), track (n = 17) or no-instructions (n = 13) group. In each group, participants had to complete a conditional discrimination task consisting of two phases. During Phase 1, they always received points for matching stimuli according to their physical similarities, while during Phase 2 points were always delivered for matching stimuli according to their physical dissimilarities. Before completing both phases, participants in the rules groups received accurate instructions about the task-contingencies of Phase 1. The no-instructions group, however, did not receive such information and as such had to learn about these contingencies via trial-and-error. The results, of the data of those participants that were included for analyses, showed that when the contingencies reversed (Phase 2), participants were generally inclined to stick to behavior that was reinforced during Phase 1. This was more so for the instruction groups compared to the no-instructions group, and for the ply compared to the track group.

[4] Broadly speaking, a ply specifies consequences delivered by the rule-giver for compliance with the rule (e.g., "I will give you money if you follow my [i.e., the experimenter] instructions"), while a track describes consequences that occur naturally when following the rule (e.g., "I will feel less pain if I take a pain-killer"). See Kissi et al. (2017) for more information on plys and tracks.

Kudadjie-Gyamfi & Rachlin (2002) examined the impact of rule-governed versus contingency shaped behavior on adaption to task-contingency changes. Eighty participants were randomly divided into an instruction (n = 40) and a no-instruction (n = 40) group. In each group, participants had to press one of two buttons (Button 1 or 2) in order to earn points and minimize the delays between consecutive trials. During Phase 1 of the task, pressing Button 2 rather than Button 1 was more effective, because this maximized point earnings while reducing the delays between consecutive trials. During Phase 2, however, these contingencies were reversed so that now pressing Button 2 rather than Button 1 was more advantageous (in terms of more points and smaller time-delays). Before beginning the task, participants in the instructions group received accurate instructions about the task-contingencies during Phase 1, while no such information was provided to the no-instructions group. Results suggested that during Phase 2, all groups were likely to continue selecting Button 1, but this tendency was higher in the instructions groups compared to the no-instructions group.

Lefrancois, Chase, and Joyce (1988) examined how receiving accurate instructions or no instructions about how to earn points differentially affected participants' adaptation to changes in reinforcement schedules. Participants were randomly assigned to one of six groups: Variety 1 instructions multiple reinforcement schedules (n = 15), Variety 2 instructions multiple reinforcement schedules (n = 15), Specific instructions Variable-Interval (VI) schedule (n = 15), Specific instructions Variable-Ratio (VR) schedule (n = 15), Minimal instructions VI schedule (n = 15) or Minimal instructions VR schedule (n = 15) group. During Phase 1 of the task, all groups except the Minimal instruction groups, received instructions which accurately described the way to earn points. In the variety instructions groups, multiple accurate instructions were given across a variety of reinforcement schedules, while in the specific instructions groups only one such instruction was provided under a single reinforcement schedule. During Phase 2, the task-contingencies were changed so that participants now had to earn points under an FI 30 schedule. The results showed that all groups did not behave in line with the novel reinforcement schedule. In fact, the Minimal instructions groups and the Specific instruction VR schedule group deviated the most from the task-contingencies (i.e., emitted more responses during the FI 30 schedule) compared to the Specific instruction VI schedule and the Variety instructions groups.

Monestès et al. (2017) examined whether rule-based insensitivity to task-contingency changes would generalize to other indirectly related and novel task-contingencies. In this study, participants were required to complete two tasks. In Task 1, they had to earn as many points as possible according to a VR 8 and a DRL 8 schedule in the presence of nonsense words A and B, respectively. During Task 2, participants had to match nonsense words according to the equivalence class in which they were being trained. Depending on the condition to which they were allocated, participants either received (n = 46) or did not receive (n = 41) any instructions about the task-contingencies in both tasks. Following completion of Tasks 1 and 2, participants were required to complete Task 3. This was largely similar to the first task, with two exceptions. First, instead of using the nonsense words A and B, other nonsense words that were in the same

equivalence classes as these words (trained in Task 2) were used. Second, the reinforcement contingencies were now reversed so that reinforcement was delivered according to a VR 8 schedule when stimuli in the same equivalence class as nonsense word B were shown, and a DRL 8 schedule when those belonging to equivalence class A were presented. The results, of the data of those participants that were included for analyses, revealed that both the instructions and no-instructions groups failed to fully adapt to the reversed task-contingencies during Task 3. However, this tendency was greater in the instructions compared to the no-instructions group.

Monestès et al. (2014) examined the impact of different types of instructions or no-instruction upon participants' reactions to changes in task-contingencies. Participants were either randomly provided with socially-generated instructions about the task-contingencies (n = 10), asked to generated their own rules about these contingencies (n = 10) or not giving any instructions about how they should respond in the task (n = 10). Next, they completed a task in which points could be initially earned for pressing a right button according to an FR 8 schedule, and a left one according to an FI 8 schedule. After a while, the initial task-contingencies were reversed such that points were now delivered according to an FR 8 schedule for left button presses and an FI 8 schedule for right button presses. The results showed that when the task-contingencies reversed, participants failed to adapt to this reversal (i.e., they continued to press the right button more frequently than the left button). This was more the case in the socially-provided rule group, followed by the self-instructed group, and then the no-instructions group.

In two experiments (Experiment 1: n = 100; Experiment 2: n = 96), Otto, Torgrud, and Holborn (1999) tested the effects of instructions on participants' adaptation to contradicting task-contingencies. Participants were required to press computer keys to move a cursor through a matrix. Points for cursor movements were initially delivered under a multiple FR 18 and a DRL 6 schedule, where each component alternated every few minutes. Before being exposed to this phase, participants received accurate instructions to go fast and slow when the FR 18 and DRL 6 schedules were in effect, respectively. After a while, the task-contingencies were reversed so that now participants were instructed to go fast when the DRL 6 and slow when the FR 18 schedules were in effect. The results showed that, in both Experiments, participants failed to adapt fully to the task-contingency change.

Shimoff, Catania, and Matthews (1981; Experiment 1) examined how instructed versus non-instructed participants adapted to task-contingency changes. In this study, participants could initially earn points by pressing a button slowly during a combined Random-Interval (RI) 15 and DRL 3 schedule. After a while, however, the reinforcement contingency during the DRL 3 schedule was removed, so that points could only be earned under the RI 15 schedule. Before initiating the experimental task, participants were either accurately informed about the task-contingencies that were in effect prior to the contingency change (but not those after this change) (n = 10) or received no such information (n = 11). Results showed that after the contingency change, both groups failed to behave in line with this change and that this effect was larger for participants that were given instructions.

Souza, Pontes, and Abreu-Rodrigues (2012) investigated the effects of changes in the accuracy of instructions to emit systematic or random digit sequences on participants' behavior. To evaluate this, Souza et al. randomly assigned participants to a systematic (n = 12) or random instructions (n = 12) group, or a group that did not receive instructions about the task-contingencies (n = 12). In each of these groups, participants completed a task in which they had to type sequences of three digits which, if correct, were always rewarded with points. During the first phase of the task, a sequence was considered correct if it a) differed from the two previous sequences and b) had a weighted relative frequency that was less than or equal to a certain threshold. During the second phase, however, this contingency was omitted and, as a result, responding no longer produced reinforcement. Results, of the data of those participants that were included for analyses, indicated that during the second phase, all participants continued to respond in ways that were effective during Phase 1. This effect was slightly more pronounced in the systematic instructions group compared to the random and no-instructions groups.

Svartdal (1989), examined how receiving instructions affected adjustment to inaccurate response-feedback. Participants completed a task in which they were told to count and correctly report the number of auditory stimuli they heard. During the first few trials, no feedback was provided about the accuracy of their reports (i.e., baseline). After a while, however, participants received feedback about their reports (i.e., during the feedback trials). Unbeknownst to participants, this feedback was not based on the accuracy of their reports but rather on the rate with which they reported the number of stimuli they heard. That is, feedback was delivered whenever participants' mean rate of responding was either below (Slow group; n=14) or above (Fast group; n =13) their baseline rate of responding. According to the authors, during the feedback trials, participants in the Slow group were slower and those in the Fast group faster to emit responses (compared to baseline), which suggests that participants generally adapted to the novel contingencies. This tendency, however, was slightly more pronounced in the Slow compared to the Fast group.

Svartdal (1995; Experiment 2) explored the impact of instructions on participants' adaption to changes in both instructed- and task-contingencies. First, participants were informed that during the first part of the task, correct responding would be reinforced with a light signal whenever they pressed a key once every second. They were then told that during Part 2, reinforcement (i.e., a light signal) would be delivered if they slightly decreased (n = 12; Decrease group)[5] or increased (n = 12; Increase group)[6] their response rate and kept this rate as stable as possible for the remainder of the experiment. Participants were additionally informed that during Part 2, feedback about their responding would be less informative and that they should, therefore, base their performances on what they had learned from Phase 1. Results showed that

[5] This number is based on an educated guess, given that the exact number of participants within each experimental group was not provided. It was merely stated that subjects were randomly assigned to one of the two experimental groups.
[6] See Footnote 5.

participants adapted to the contingency change, given that during Phase 2 rates of responding declined and augmented in the Decrease and Increase groups, respectively.

Torgrud et al. (2006; Experiment 1) examined how initially accurate instructions on either a functional or non-functional multiple reinforcement schedule, or a VR 8 schedule differently impacted participant responding on an FI 30 schedule. All participants were randomly assigned to one of three groups: the functional multiple (n = 15), non-functional multiple (n = 15) or single (n = 15) schedule group. In each of these groups, participants were instructed to try to earn as many points as possible in order to increase their chances of winning a monetary reward. In the multiple schedule groups, participants initially received instructions which accurately informed them about how they could earn points by pressing an "earn" key under an FR, a DRL, and a VI schedule. These contingencies could either be functional or non-functional depending on whether they trained a response rate that was or was not beneficial under the FI 30 schedule, respectively. Participants in the single schedule group, however, only received accurate instructions which initially informed them about how they could earn points under a VR 8 schedule. After some trials, all participants were then exposed to the FI 30 schedule. The findings showed that all groups failed to adapt to the last two minutes of this schedule. This was more pronounced in the single schedule group compared to the other groups, and in the non-functional schedule group compared to the functional schedule group.

Torgrud et al. (2006) attempted to replicate and extend these findings in a second experiment in which 150 participants were randomly assigned to one of six multiple schedule groups: Functional FR (F-FR), Non-Functional FR (NF-FR), Functional DRL (F-DRL), Non-Functional DRL (NF-DRL), Functional FR and DRL (F-BOTH) or Non-Functional FR and DRL (NF-BOTH), or a single schedule group. As in the previous experiment, participants initially received accurate instructions before being exposed to an FI schedule (now an FI 15 as opposed to an FI 30). This time, these instructions described how participants could earn points during an FR, a DRL, a VI, a tandem DRL, and a tandem VI schedule in the multiple schedules groups, and a VR 8 schedule in the single schedule group. Once again, the functionality of these reinforcement contingencies depended on the extent to which they were useful to gain points under the FI 15 schedule. The results showed that overall, all groups failed to adapt to the task-contingencies during the last two minutes of the FI 15 schedule and that this was more prominent in the single schedule group, followed by the FR, the BOTH, and the DRL schedule groups, in that order.

Studies included to answer Research Question 2. Of those studies that met our inclusion criteria, only Baruch et al. (2007) examined whether psychological suffering moderated the RBIE. Specifically, Baruch et al. examined whether different types of instructions (plys and tracks) and the presence or absence of sub-clinical symptoms of depression differentially impacted adaption to task-contingency changes. Non-depressed (n = 14) and depressed (n = 15) undergraduate students were randomly given a ply or track which initially correctly described the task contingencies in a matching-to-sample (MTS) task, but in a

subsequent phase were in contrast with these contingencies. The results revealed that both groups showed difficulties adapting to the new task-contingencies. However, relative to the non-depressed group, the depressed group adapted quicker to this change. No differences were observed as a function of the plys or tracks these groups received.

Appendix S2:

Judgement of the relevant domains of risks of bias

Studies used to answer Research Question 1 ("Is there evidence for the rule-based insensitivity effect in adults?").

	Non-random sequence generation **(Selection Bias)**	Allocation revelation **(Selection Bias)**	Prior testing **(Selection Bias)**	Misclassification of participants to experimental groups **(Selection Bias)**	Incomplete outcome data **(Exclusion Bias)**	Selective reporting of outcomes **(Reporting Bias)**	Invalid and unreliable outcome assessment methods **(Detection Bias)**
Cerutti (1991)	?	?	?	NA	?	-	-
Cerutti (1994)	?	?	?	NA	?	-	-
Dixon et al. (2000)	?	?	?	+	?	-	-
Haas and Hayes (2006)	?	?	?	+	?	-	-
Harte et al. (2017 - Experiment 1)	?	?	?	NA	?	-	-

Harte et al. (2017 - Experiment 2)	?	?	?	+	?	-	-
Hayes et al. (1986)	?	?	?	+	?	-	-
Kissi et al. (2018)	-	?	?	+	?	-	-
Kudadjie-Gyamfi and Rachlin (2002)	?	?	?	+	?	-	-
LeFrancois et al. (1988)	?	?	?	+	?	-	-
Monestès et al. (2017)	?	?	?	+	?	-	-
Monestès et al. (2014)	?	?	?	+	?	-	-
Otto et al. (1999 - Experiment 1)	?	?	?	NA	?	-	-
Otto et al. (1999 - Experiment 2)	?	?	?	NA	?	-	-

1025

Shimoff et al. (1981)	?	?	?	+	?	-	-
Souza et al. (2012)	?	?	?	+	?	-	-
Svartdal (1989)	?	?	?	NA	?	-	-
Svartdal (1995 - Experiment 2)	?	?	?	NA	-	-	-
Torgrud et al. (2006 - Experiment 1)	?	?	?	NA	?	-	-
Torgrud et al. (2006 - Experiment 2)	?	?	?	NA	?	-	-

1026

	Inadequate outcome assessments **(Detection Bias)**	Inadequateness of the method used to determine sample size **(Detection Bias)**	Inappropriateness of analytic methods **(Detection Bias)**	Non-standardization of the experimental context **(Performance Bias)**	Information about the study objectives (**Performance bias**)	Non-Blinding of participants and personnel (**Performance bias**)
Cerutti (1991)	-	?	?	-	?	?
Cerutti (1994)	-	?	?	-	?	?
Dixon et al. (2000)	-	?	?	-	?	?
Haas and Hayes (2006)	-	?	?	-	?	?
Harte et al. (2017 - Experiment 1)	-	?	?	-	?	?
Harte et al. (2017 - Experiment 2)	-	?	?	-	?	?
Hayes et al. (1986)	-	?	?	-	?	?
Kissi et al. (2018)	-	?	?	-	?	?

1027

Kudadjie-Gyamfi and Rachlin (2002)	-	?	?	-	?	?
LeFrancois et al. (1988)	-	?	?	-	?	?
Monestès et al. (2017)	-	?	?	-	?	?
Monestès et al. (2014)	-	?	?	-	?	?
Otto et al. (1999 - Experiment 1)	-	?	?	-	?	?
Otto et al. (1999 - Experiment 2)	-	?	?	-	?	?
Shimoff et al. (1981)	-	?	?	-	?	?
Souza et al. (2012)	-	?	?	-	?	?
Svartdal (1989)	-	?	?	-	?	?

1028

Svartdal (1995 - Experiment 2)	-	?	?	-	?	?
Torgrud et al. (2006 - Experiment 1)	-	?	?	-	?	?
Torgrud et al. (2006 - Experiment 2)	-	?	?	-	?	?

Note. '+', '-', and '?' refer to high, low, and unclear risk of bias for a particular domain, respectively. NA means that the domain was not applicable.

Appendix S3:

Judgement of the relevant domains of risks of bias

Studies used to answer Research Question 2 ("Do adults suffering from psychological problems display a larger RBIE compared to their non-clinical counterparts?").							
	Non-random sequence generation **(Selection Bias)**	Allocation revelation **(Selection Bias)**	Prior testing **(Selection Bias)**	Misclassification of participants to experimental groups **(Selection Bias)**	Incomplete outcome data **(Exclusion Bias)**	Selective reporting of outcomes **(Reporting Bias)**	Invalid and unreliable outcome assessment methods **(Detection Bias)**
Baruch et al. (2007)	?	?	?	NA	?	-	-
	Inadequate outcome assessments **(Detection Bias)**	Inadequateness of the method used to determine sample size **(Detection Bias)**	Inappropriateness of analytic methods **(Detection Bias)**	Non-standardization of the experimental context **(Performance Bias)**	Information about the study objectives (**Performance bias**)	Non-Blinding of participants and personnel (**Performance bias**)	
Baruch et al. (2007)	-	?	?	-	?	?	

Note. '+', '-', and '?' refer to high, low, and unclear risk of bias for a particular domain, respectively. NA means that the domain was not applicable