Patterns of Relational Framing in Executive Function: An Investigation of the Wisconsin Card Sorting Test

Mårten J Tyrberg^{1,2} • Thomas Parling³ • Tobias Lundgren³

Accepted: 4 February 2021 / Published online: 15 March 2021
© The Author(s) 2021

Abstract

Executive function (EF) is a concept widely used to describe higher order cognitive functioning. However, the concept seems to lack a precise definition upon which even experts in the area can agree. Examples of existing definitions describe EF as capacities for independent self-serving behavior and aspects of working memory, planning and organizational skills. It has been argued that the definition of EF is too imprecise, in terms of informing effective interventions. To possibly understand the range of behaviors termed EF in a more workable way, we administered a well-known computerized test (the Wisconsin Card Sorting Test; WCST) of EF to a sample of normal functioning participants ($n = 11$). During administration, they were asked questions about their ways of thinking. Transcriptions of questions and answers were analyzed in search of common patterns, using a behavioral theory of language and cognition (relational frame theory; RFT). The analysis suggested that a specific relationship between deictic and temporal framing, on the one hand, and coordination and spatial framing, on the other, was in effect when participants performed well on the WCST. At critical time points during the test (i.e., when need for EF was arguably most acute), deictic and temporal framing dominated participants' behavior, whereas coordination and spatial framing were used relatively less.

Keywords Executive function · Relational frame theory · Wisconsin Card Sorting Test · Behavior therapy

In an article dealing with the central tenets of contextual behavioral science, Wilson (2016) makes a statement that is deceptively simple: “Terms are merely ways of speaking” (p. 62). Indeed, they are. And yet terms, with time, come to carry a lot of weight. This makes it increasingly difficult to know what is meant when certain terms are used. It also makes it difficult to hold terms lightly. Executive function (EF) is a term that carries a lot of weight, in the sense that it is used frequently in clinical psychiatric practice, as an important domain of cognitive functioning. At the same time, when one tries to find a clear definition of what the term refers to more

precisely, it proves difficult. The explanatory value of the term EF has been argued to be low (Hayes et al., 1996).

A few sample definitions of EF out of many are “capacities that enable a person to engage successfully in independent, purposive, self-serving behavior” (Lezak, 1995, p. 42), “processes that are responsible for guiding, directing, and managing cognitive, emotional, and behavioral functions” (Gioia et al., 2000, p. 1) and “a variety of different capacities that enable purposeful, goal-directed behavior, including behavioral regulation, working memory, planning and organizational skills, and self-monitoring” (Mangeot et al., 2002, p. 272). Barkley (2012) refers to a scientific conference where 10 experts in neuropsychology were asked to generate terms considered to be EF and came up with 33 such terms. Indeed, there seems to be a consensus even among EF knowledgeable researchers that the term lacks an acceptable operationalization.

Despite the apparent lack of precision, there are several standardized tests designed to measure EF. The Wisconsin Card Sorting Test (WCST; Heaton et al., 1993) is one such test commonly used in clinical psychiatric practice. It was originally designed as a test for “abstract reasoning ability and the ability to shift cognitive strategies in response to changing environmental contingencies” (Heaton et al., 1993, p. 1). The test has been standardized and normed for persons

✉ Mårten J Tyrberg
marten.tyrberg@regionvastmanland.se

¹ Uppsala University, Uppsala, Sweden

² Centre for Clinical Research, Hospital of Västmanland, Region Västmanland, Västerås SE-72189, Sweden

³ Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Region Stockholm, Norra Stationsgatan 69, Stockholm SE-113 64, Sweden

aged between 6½ through 89 years of age. Studies (Perrine, 1993; Shute & Huertas, 1990) suggest that the WCST, compared to other similar tests, measures distinct aspects of frontal lobe function and this has been taken (Heaton et al., 1993) as support for the construct validity of WCST as a measure of EF.

As a clinical tool, the WCST has been proven as a predictor of various relevant outcomes. For example, Wicks et al. (2001) reported that in an inpatient sample of patients with alcohol dependence, there was a significant correlation (large in terms of effect size) between WCST performance and subsequent nondrinking days. Lysaker et al. (2005) found, in a group of schizophrenia patients, that WCST performance predicted work performance, with effect sizes in the moderate range. Further, assessing the ecological validity of the WCST in a nonclinical sample, Kibby et al. (1998) found that it correlated significantly with occupational status or type of position held by an individual.

The WCST is a task given with few instructions. In the computerized version, the subject is presented with four stimulus cards at the top of the screen, each with different stimulus features according to number (1–4), shape (triangle, star, cross, or circle) and color (red, green, yellow, or blue). At the bottom of the screen, a stack of cards is presented, each with varying stimulus features according to the same system. The task is to sort these cards beneath either of the four stimulus cards. Instructions consist of telling the subject to try to sort the cards correctly, and that the computer will give feedback on whether the sorting is correct or incorrect. No other instructions are given. After each trial, the computer feedback is either “correct” or “incorrect.” At the beginning of the test, a specific correct sorting rule (color) is in effect. After a fixed number of correct consecutive trials, the sorting rule is changed, without the subject’s knowledge. The test ends after the subject has completed a certain number of categories, or after 128 trials if all categories are not completed.

Relational frame theory (RFT; Hayes, Barnes-Holmes, et al., 2001) is a theory of language and cognition, based on learning theory (i.e., relying on principles of operant conditioning). At its core is a class of behavior termed arbitrarily applicable relational responding (Hughes & Barnes-Holmes, 2016a). This refers to how verbally able humans relate stimuli not only based on formal properties (such as physical size) but also on arbitrary relations (such as value). An example of such a relation is that between two coins, where one is physically smaller but worth more than the other. The relation based on value is arbitrary, whereas the relation based on formal properties is not.

In RFT, an event is defined as verbal if it shows the contextually controlled qualities of mutual entailment, combinatorial entailment and transformation of stimulus functions (Barnes-Holmes et al., 2001). If these criteria are fulfilled, the event is said to participate in relational frames. According to RFT, a typically developing human being learns

to relate stimuli in increasingly complex ways (Hayes, Fox, et al., 2001; Hughes & Barnes-Holmes, 2016b). Though not a developmentally fixed sequence, in general the first step is responding to events based on frames of coordination (e.g., the spoken word “dog” is the same as an actual dog), followed by frames of opposition (e.g., “good” is the opposite of “bad”), distinction (e.g., cats are different from dogs), comparison (e.g., someone being “older than” someone else), hierarchy (e.g., something or someone belonging to a category, group, or class), temporality (e.g., summer comes after spring), spatiality (e.g., an object being “over there”), conditionality/causality (i.e., events being related so that causality is implied, e.g., being hit “causes” pain) and deictics (e.g., relating from the perspective of “I”).

The first, and to our knowledge so far only, explicit discussion of EF from an RFT perspective (Hayes et al., 1996) proposes that EF is a subset of rule-governed behavior. EF tests such as the WCST, thus, aim to investigate the conditions under which “people select among available rules or generate new ones, follow rules when they are available even though they conflict with other sources of behavioral control, and change them when they no longer work” (Hayes et al., 1996, pp. 292–293). In some sense, the tests train rules and examine their flexibility when they no longer apply. In the WCST, specific contextual cues provide consequences (the computer response “incorrect”). Hayes et al. (1996) also argue that EF are not primarily responses that are well practiced or automatic. They typically come into play when the subject is met by contextual features that signal the introduction of something new, something that has not been encountered before. On such an occasion, there will be nothing in the formal properties of the context that tells the subject what response to make next to be reinforced (i.e., to produce the computer response “correct”).

Rule-governed behavior, as defined by RFT, is a learned behavior controlled by the correspondence between relations specified in a certain rule, and the behavior emitted by a person (Hughes & Barnes-Holmes, 2016b). In colloquial terms, it is behavior reinforced by doing what is right or correct (according to the rule). Rules might be described as relational networks (Hughes & Barnes-Holmes, 2016b), within which transformations of stimulus functions occur, depending on which relations are specified. For example, during WCST performance, a workable rule might be: “Sorting according to color was correct just now, so I will try that again.” In this rule, “sorting according to color” is in a frame of coordination with “correct.” There also seems to be a conditional/causal frame relating previous success (“was correct just now”) with possible future behavior (“will try that again”). I-again indicate deictic framing (relating from the perspective of “I”), and there are also cues for temporal framing (now, again). Hence, from an RFT point of view, the flexibility of rule-governed behavior (upon which EF presumably depend [Hayes et al.,

1996]) rests on the nature of relations between stimuli (e.g., coordination), and the functions that are involved (e.g., being correct).

From an RFT point of view, there seems to be no way of demonstrating a truly nonverbal act in a verbally able human being (Hughes et al., 2012). Therefore, in the present study, we focus on *how* the class of behaviors emitted by our subjects during WCST performance is verbal. That is, what characterizes this specific type of verbal behavior? In colloquial terms, the behavior we are interested in is “thinking.” From the perspective of RFT, thinking is defined as “a reflective behavioral sequence, often private, of pragmatic verbal analysis that transforms the functions of the environment so as to lead to novel, productive acts” (Hayes, Gifford, et al., 2001, p. 95). Pragmatic verbal analysis (PVA), in turn, has been defined as “conceptualizing and verbally manipulating aspects of our non-arbitrary environment so that we may respond to that environment more effectively” (Stewart et al., 2013, p. 176). PVA might be said to be the RFT conceptualization of what is colloquially termed problem solving.

We realize that it might be considered contradictory to first argue for the lack of an agreed upon definition of EF, and then accept that a measure such as the WCST in fact measures EF. It is important to note that in choosing WCST performance as a measure of EF, we do not claim that it captures an underlying real entity. Based on its common use in clinical practice, and its potential to predict clinically relevant outcomes, we claim that the WCST is a reasonable approximation of the wide range of behaviors that are, in many current definitions, included in the concept of EF.

Further, we do not claim that the verbal statements that are subjects of analysis in the present research correspond in a one-to-one manner to the actual thinking they describe (presumably EF). We do argue, however, that asking people what the contents of their thoughts are is a reasonable approximation of observing the actual thoughts. A behavioral protocol for letting people talk aloud during the performance of tasks, and apply controls for whether what they say out loud functionally corresponds to self-rules that govern their behavior (i.e., “thinking”) has been formulated by Hayes et al. (1998). It has been termed the “silent dog” method. As will be made clearer below, our approach differs from the “silent dog” method in the sense that we ask specific questions to our subjects at prespecified time points, based on assumptions regarding when EF is at play. We also do not have experimental control over the contextual conditions of subjects’ actual thinking. Rather, we infer function from topography of language.

Previous theoretical research regarding EF affords a central role to rule-governed behavior. So called rules, in turn, have been conceptualized as consisting of relational networks, within which transformations of stimulus functions occur in accordance with different families of relational framing (e.g.,

coordination, temporal). Although interested in function, we argue that a group-level approach is a suitable first exploratory step, because at this stage of what one hopes is longer research journey we are primarily interested in overarching patterns common across participants.

The first aim of the present study is to test whether it is feasible to collect verbal statements about thinking and code these in a systematic manner, thus making it possible to track common patterns of families of relational framing in the context of a specific kind of problem solving (WCST performance as an approximation of EF).

The second aim is to explore which verbal behaviors, in particular the distribution of relational frames, are involved in the phenomenon referred to as EF (as it is measured by WCST) and to generate hypotheses about which patterns of relational framing constitute the phenomenon. The WCST has been chosen as a standardized operationalization of the umbrella term EF, based on its common use in clinical practice.

Our research questions are:

- Is it feasible to develop a system for coding how people speak about their thoughts, and yield meaningful descriptions of purported underlying verbal behavior in terms of families of relational framing?
- On a group level, what are the patterns of relational responding common among participants when they solve the problems of a typical test of EF (WCST)?
- What is the relationship between observed group level patterns of relational framing during WCST performance and test outcome variables?

Method

Design

The study was an observational study using a combination of quantitative and qualitative methodology.

Participants

Participants were recruited among staff working in the psychiatric clinic of the Hospital of Västmanland, Västerås, Sweden. Posters were placed in staff recreational areas, encouraging staff members to contact the principal researcher (first author) for inclusion in the study. The posters also included brief information about the purpose of the study, how long testing would take, and what compensation participants would receive. Participants were rewarded with a gift certificate at a movie theatre, corresponding to the entry fee for two persons. They could participate in the study during working hours. Recruitment and testing took place between May and

June 2019. When consenting to participate (or at request) participants also received more thorough written information about the purpose of the study, the ethical obligations of the researchers, and contact information for all responsible researchers. Demographic characteristics of participants are displayed in Table 1. None of the participants suffered from a severe somatic condition.

Materials

In the standard administration of the WCST (Heaton et al., 1993), which can either be with physical cards or computerized, a minimum of 60 and a maximum of 128 trials are administered. The correct sorting rule (category) in effect is changed after the subject has made 10 consecutive correct responses. The sequence of categories is color, shape, number, color, shape, and number. Thus, each category is repeated once. The test ends when the subject has completed six categories (which requires a minimum of 60 correct responses) or when the maximum number of trials has been administered. Thus, there is a maximum of five changes of the correct sorting rule in effect.

The outcome variables used for analysis in the present study, explained below, were all drawn from the original WCST test manual (Heaton et al., 1993) and are displayed in the protocol that is generated after a standard testing procedure (except for the composite score). In the analysis of WCST results, age and education demographically corrected norms were used. Five of the outcome variables of the WCST are standardized ($M = 100$, $SD = 15$), allowing comparisons among individuals and various patient groups. Higher scores correspond to better performance. *Total errors* reflect the total number of incorrect responses made during the test. *Perseverative responses* refer to when the participant persists in responding to a stimulus characteristic that is incorrect, for example color. A perseverated-to principle is established the first time a participant makes an unambiguous incorrect response. The most common situation when this will occur is when the correct sorting principle changes. *Perseverative errors* are responses that matches a perseverated-to principle, and that are incorrect (because sorted cards might match stimulus cards on more than one stimulus characteristic, perseverative responses might on occasion be correct).

Nonperserverative errors are incorrect responses that do not match a perseverated-to principle. Finally, *percentage conceptual level responses* are defined as three or more consecutive correct responses. The term “conceptual” refers to the presumption that to make three consecutive correct responses some insight into the correct sorting category is required. *Learning to learn* is an additional outcome variable of interest and reflects a participant’s “average change in conceptual efficiency across the consecutive categories (stages) of the WCST” (Heaton et al., 1993, p. 13). Positive learning to learn scores indicate improved efficiency across consecutive categories.

Procedure

The computerized version of the WCST was administered to all participants. Testing took place in the first author’s office, except in one case where testing took place in a separate room for psychological testing. Administration of the test, including questions, took an average of 20 min and 58 s ($SD = 10$ min 26 s). Participants were seated by a desk, with a laptop computer with a 14-in screen. They were given the choice of using the desk keys or the computer mouse to manage the test. The test administrator (first author) was seated next to the participants throughout the testing. Before testing commenced, the participants were given standardized information about test procedures (Heaton et al., 1993). In addition, they were informed that during testing the administrator would ask questions, which would be audio recorded. Aside from the verbal information beforehand that questions would be asked during testing, the fact that testing was audio recorded and the fact that questions were in fact asked during testing, the administration of the WCST in the present study did not differ from how it is described in the manual (Heaton et al., 1993), or how it is usually administered in clinical practice. During testing, all participants were asked the same set of questions, at specified points of time (see Table 2 for a description). In addition, follow-up questions were asked as needed.

Each testing was audio recorded in full, from the moment after the standardized instruction had been given until the testing was completed, using the laptop computer’s built-in microphone. After testing all participants, each sound recording was transcribed verbatim by the first author.

Data Analysis

RFT analysis The method of analysis, in terms of identifying relational framing patterns in transcribed text, was established through joint discussion among the three authors. Using nine principal families of framing (Hayes, Fox, et al., 2001) as a template (this taxonomy is somewhat arbitrary, others exist in the RFT literature [e.g., Hughes & Barnes-Holmes, 2016a]), topographical cues (i.e., words) thought to indicate the use of

Table 1 Participant ($n = 11$) Characteristics

Mean age (SD)	32.27 (9.63)
Gender	
Female	9
Male	2
Mean years of education (SD)	15.00 (1.07)

Table 2 Questions and structure of questioning

Question	Point of time				
	Before starting	Incorrect response directly after category shift	Incorrect response, irrespective of point of time	First correct response after category shift	Second correct response after category shift
What do you think you can do now?	x	x	x		
How come you think so?	x	x	x		
Why do you think that was incorrect?		x	x		
Why do you think that was correct?				x	x

the respective type of framing were generated. The cues exemplified in a methodologically similar study (Belisle et al., 2018) were translated to Swedish and used as a starting point. Additional cues were generated through discussion among all three authors. Through repeated reading of all 11 transcribed texts, additional cues were added. The final list of topographical cues is presented in Table 3, together with the nine families of framing and a brief description of each. Note that Swedish cues have been translated directly into English, so that all corresponding word forms in English might not be represented. As an example of the process, the cues chosen for hierarchical framing were those words used by participants when the underlying thought process was inferred to include hierarchical framing (e.g., a participant stating “I could either choose color or shape”—indicating framing color and shape

as belonging to the class possibly correct responses, see, e.g., Stewart et al., 2018 for a discussion on hierarchical framing).

In the analysis of the transcribed texts, the search and replace function in Microsoft Word was used. Each family of framing was assigned a color. For example, the search and replace function was set to replace the cue “not” with the cue “not” highlighted in blue. Each transcribed text was searched for all topographical cues in Table 3. Searching and replacing was done by hand, in the sense that each time a topographical cue indicating the use of a type of framing (e.g., “not” for distinction framing) appeared, the analyst assessed the context of the cue, and decided whether it should be highlighted or not. That is, an idiographic functional assessment was conducted for each individual case (i.e., does this topographical cue seem to indicate the use of this type of

Table 3 Families of relational framing, and topographical cues indicating each one

Family of framing	Description (Hayes, Fox, et al., 2001)	Topographical cues
Coordination	Refers to responding in terms of identity, sameness or similarity.	Is, same, like, was, become, became, agree, correspond, goes with, fits.
Opposition	Indicates that two events are at about the same, but opposing, distance from a reference point on a continuum.	Opposite, contrary.
Distinction	Refers to responding to something being different from something else.	Not, unlike, or, change, other, different, instead, change, grow into, but, none.
Comparison	Where “one event is responded to in terms of a quantitative or qualitative relation along a specified dimension with another event” (p. 36).	Better, worse, more, less, fewer, under, over, most, least, best, worst.
Hierarchical	Refers to responding in terms of classes, groups, etc.	Or, alternative, either, choice, choose, possibility.
Temporal	Responding that arranges events along a temporal continuum.	Then, now, later, before, earlier, after, when, again, last, just now, first, next, there, still, yet, at the same time, this time, already, longer, beginning, recently, ahead, once again.
Spatial	Refers to responding in terms of how events are arranged in space.	Here, there, side, in front of, behind, over, under, by, above, where, on.
Conditionality/causality	Responding where events are arranged so that conditionality/causality is implied.	If, so, then, because, become, when, became, because, in that case, since, because of, depend.
Deictic	Responding that specifies “a relation in terms of the perspective of the speaker” (p. 38).	I, we, you, one, me.

framing in this case?). For example, the cue “is” in some cases indicated coordination framing, but not in cases where it was followed by the word “not.” “I” in some cases indicated deictic framing, but only in cases where there was also a context of temporal or spatial framing or an otherwise clear shift in perspective (e.g., “we think in different ways, I and this computer”). In cases where the same cue was assessed to indicate use of two types of framing, it was highlighted in one color and the text was assigned another color (e.g., “or” in some cases indicating both distinction and hierarchical framing).

Analyzing instances of comparison framing required manually searching each text, because it relied on cues that could be any comparative or superlative form of an adjective. Thus, for comparison framing, each text was read through separately, in addition to the search for the cues in Table 3.

All three authors analyzed the first part of the first transcribed text to establish consensus regarding the method of analysis. The remaining analyses were conducted by the first author. After completion, all the analyzed texts were scrutinized for accuracy of analysis by the second author. During this phase, suggestions for additional highlights were made, predominantly within the deictic family (an additional 93 deictic, 3 coordination, 2 hierarchy, and 1 distinction highlights were added). Some additional topographical cues were added during this phase as well (“goes with” and “fits” for coordination, “none” for distinction, and “possibility” for hierarchy). All texts were searched for each newly added topographical cue.

Results from the RFT analysis were quantified by counting the number of highlights from each family of framing for each respective text. During this phase, the family of opposition framing was removed from analysis, because no cues belonging to this family had been identified.

In the next step, highlights were categorized according to where in the testing procedure they appeared. Six categories were used, namely the first screen, incorrect response directly after category shift, incorrect response irrespective of when, correct response directly after category shift, correct response irrespective of when, or other. Most highlights were categorized in either of the first five categories. The category used in the analyses of results was incorrect response directly after category shift (from here on referred to as category shift). It can be argued that when the test calls for the generation of a new sorting rule, as is required at category shift, the need for EF is the most acute. This category included all statements made between the point where the computer program changed the correct sorting rule, and to the point where the participant made a correct response according to the new sorting rule. The absolute number of highlights from each respective family of framing was transformed into relative numbers (percentages), to adjust for the variation in length between the various transcribed texts. Percentages were produced by dividing the absolute number of highlights from each family with the sum of

highlights from all families for each category and multiplying the result with 100. Percentages were used in all statistical analyses.

Statistical analysis In the first set of analyses, because some variables were found to be nonnormally distributed according to Shapiro Wilk tests of normality, nonparametrical Wilcoxon Signed-Rank Tests were used to analyze differences between patterns of framing at different time points. The variables in the second set of analyses were all found to be normally distributed. Therefore, Pearson correlation analyses were used to investigate correlations between patterns of framing and WCST outcomes.

Results

Feasibility

Our first research question concerned whether the development of a coding system was at all feasible. We found that it was reasonably feasible to develop an RFT-based system for coding how people speak about their thoughts. However, as it was constructed in the present study, it also demanded a good deal of discussion before reaching a reasonable consensus. It was a wholly exploratory process, relying on interpretation and a dynamic and ongoing adjustment of our tools for coding, in the sense that the tools were fine-tuned while coding itself was in progress. The development of the coding system relied on continuous back and forth discussion among all three authors, moving from actual data (the transcribed texts) to overarching theory (e.g., discussing definitions of principal families of framing).

Patterns of Relational Responding

On a purely descriptive level, the average pattern of relational framing was dominated by coordination framing, both during the whole test, and in particular at category shift. In both cases, comparison framing was used the least, relatively speaking. Figure 1 displays the patterns of framing for the respective families, during the whole test and in particular at category shifts. For the reader to be able to see for each subject which families of framing dominated during different points in time, and possible individual connections to correct responding, Table 4 displays the patterns of framing subject by subject, and the respective raw score outcomes for each.

The first step in the analysis was to explore potential differences between patterns of framing during the whole test, and patterns of framing in particular at category shift. Wilcoxon Signed-Rank Tests indicated that there was a significant difference in four families of framing, namely coordination framing (whole test $Mdn = 29.1$; category shift $Mdn =$

Fig. 1 The relative mean use ($n = 11$) of relational frames (%) during the whole test and at category shifts

26.2), $z = -2.401$, $p = .016$, spatial framing (whole test Mdn = 7.3; category shift Mdn = 5.7), $z = -2.135$, $p = .033$, temporal framing (whole test Mdn = 15.6; category shift Mdn = 20.0), $z = -1.956$, $p = .050$, and deictic framing (whole test Mdn = 9.8; category shift Mdn = 14.0), $z = -2.578$, $p = .010$. Coordination and spatial framing were used relatively less after category shift compared to the whole test, whereas deictic and temporal framing were used relatively more.

Correlations with Outcomes

To explore the patterns further, we ran a correlation analysis (Pearson) between the relative use of coordination, spatial, temporal, and deictic framing at category shift, and the five standardized outcome variables of the WCST, namely total errors, perseverative responses, perseverative errors, nonperseverative errors, and percentage conceptual level responses, plus the learning to learn variable. We also constructed a composite WCST score, by transforming all six variables into z -scores ($M = 0$, $SD = 1$), adding them, and calculating the mean. Our prediction, based on the pattern of framing revealed so far, was positive correlations for temporal and deictic framing, and negative correlations for coordination and spatial framing. The results of the analysis are displayed in Table 5. Note that all WCST outcome variables are constructed so that higher scores represent better performance.

There were significant negative correlations between coordination framing and perseverative responses as well as perseverative errors and composite WCST score. In addition, there were significant negative correlations between spatial framing and both nonperseverative errors and percentage conceptual level responses. Further, there was a clear pattern among the nonsignificant correlations; coordination and spatial framing had a negative association (except for spatial

framing and perseverative responses) whereas temporal and deictic framing had a positive association with the outcome variables of the WCST.

In Table 6, example reasoning from each participant is displayed, to illustrate how deictic and temporal framing might have been used to generate new sorting rules. All examples were drawn from the context of incorrect response directly after category shift.

Discussion

In terms of the overall pattern of relational responding, four families of framing distinguished themselves in the present study: coordination, spatial, temporal, and deictic framing. In quantitative terms, coordination was the dominant family of framing both during whole test performance, and during category shift, accounting for 29.5% and 24.3% of all identified instances of framing, respectively. This seems logical, seeing how coordination is at the basis of much of our everyday language. Referring to what something *is*, for example, relies on coordination framing. Just looking at how we typically use language to describe or talk about things, it seems natural that there will be much coordination framing in any verbal statement. The interesting bit of information that appeared about coordination framing in these data, however, was that it was significantly *less* dominant during WCST performance as an approximation of critical EF activity (i.e., at category shift) compared to the whole test. What might this mean? We argue that it does not mean that coordination framing is unimportant for EF. We think that coordination framing is important for all sorts of thinking. But it does not seem to be the family of framing that distinguishes EF from other activities of thinking.

Table 4 Subject by subject patterns of relational responding and respective WCST raw outcomes

Subject	WCST raw score outcome		Percentual distribution of respective family of frames used, during the whole test and at category shift (bold text)									
	Trials administered	Trials correct	Coordination	Distinction	Comparison	Hierarchical	Temporal	Spatial	Conditional/causal	Deictic		
1	87	69	24,4	23,1	0,6	2,1	1,7	19,5	17,3	5,4	4,6	19,1
2	77	67	33,0	15,5	0,0	9,3	12,6	10,3	9,0	7,7	6,3	12,6
3	106	75	36,1	19,2	0,0	3,3	1,7	15,0	20,5	7,3	7,7	10,4
4	74	65	43,9	9,8	1,2	0,0	0,0	13,4	21,9	3,7	0,0	18,3
5	88	73	41,9	16,9	0,0	0,6	1,1	15,6	19,6	1,9	0,0	15,0
6	96	78	22,7	11,9	0,0	1,8	0,0	12,9	26,2	14,7	6,2	22,7
7	70	63	17,1	19,4	0,9	8,8	7,7	17,1	20,5	6,0	2,6	15,7
8	71	61	29,4	17,1	0,5	5,5	4,8	17,1	20,0	9,2	5,7	12,0
9	118	91	29,1	13,7	0,8	0,5	0,0	14,2	7,4	11,6	12,8	13,7
10	93	72	17,8	20,9	0,9	2,4	2,5	16,7	19,3	13,4	14,9	11,8
11	70	63	28,9	17,3	1,9	5,8	0,0	21,2	35,5	1,9	0,0	9,6
Mean (SD)	86,4 (15,9)	70,6 (8,7)	29,5 (8,8)	16,8 (3,9)	0,8 (0,8)	3,6 (3,3)	2,9 (4,0)	15,7 (3,1)	19,7 (7,5)	7,5 (4,4)	5,5 (5,0)	14,6 (5,5)
												11,4 (3,3)
												14,6 (4,3)

Spatial framing was an uncommon family of framing in the present quantitative data. This was so for both whole test performance and at category shift, with spatial framing accounting for 7.5% and 5.5% of all instances of framing, respectively. Whereas being uncommon compared to, for example, coordination framing, spatial framing shared with coordination framing the feature of being significantly less used during supposedly critical EF activity (category shifts during WCST administration) compared to the whole test. Spatial framing did not dominate the verbal behavior of these participants but finding a workable new sorting rule seemed to rely on spatial framing being even less dominant.

Whereas spatial and coordination framing had a less dominant role during EF thinking as required by the WCST, two other families showed the opposite pattern. Temporal framing accounted for an average of 15.7% and 19.7% of all instances of framing during the whole test and at category shift, respectively, the difference being statistically significant. This family of framing seemed to be needed especially during critical EF activity as required by the WCST. So did deictic framing, accounting for 11.4% and 14.6% of all instances of framing during the test as a whole and at category shift, respectively, with the difference being statistically significant.

All in all, it seemed, for these participants, that finding a new workable sorting rule (e.g., after category shift) relied on using deictic and temporal framing over coordination and spatial framing.

In terms of how this pattern might be associated with objective outcome variables of the WCST, several of the correlations in Table 5 are negligible in size, and most are nonsignificant at the conventional level. Still, in the context of exploring data for a possible pattern, the direction of correlations in this sample seems clear, and all coefficients for the composite score are moderate or large (Cohen, 1988). Based on these data, however, we are naturally cautious in drawing conclusions for the population. It is likely that the nonsignificance of correlations was partly due to our small sample size, which was a limitation.

Although being seemingly dominant in all contexts of the test, successful performance on the WCST seemed to rely on coordination framing stepping back in favor of other families. This notion was supported by the direction of correlations between the relative use of coordination framing and outcome variables of the WCST. The less coordination framing at category shift, the better. The same was by and large true for spatial framing.

The importance of temporal and deictic framing was confirmed by the directions of correlations between their use during category shift and outcome variables of the WCST. The pattern was clear in showing that the more deictic and temporal framing, the better the outcome. This was true for all outcome variables, including the efficiency variable (learning to learn).

Table 5 Two-tailed pearson product moment correlations ($n = 11$) between seemingly central families of framing and outcome variables of the WCST

	Coordination r (p)	Spatial r (p)	Temporal r (p)	Deictic r (p)
Total errors	-.528 (.095)	-.552 (.078)	.412 (.208)	.296 (.376)
Perseverative responses	-.681* (.021)	.011 (.974)	.228 (.500)	.477 (.138)
Perseverative errors	-.714* (.013)	-.063 (.855)	.269 (.424)	.491 (.125)
Nonperserverative errors	-.357 (.280)	-.687* (.019)	.443 (.172)	.152 (.656)
Percentage conceptual level responses	-.366 (.269)	-.613* (.045)	.392 (.233)	.308 (.358)
Learning to learn	-.135 (.693)	-.057 (.868)	.197 (.561)	.474 (.140)
Composite WCST score	-.614* (.045)	-.433 (.184)	.428 (.189)	.485 (.130)

* $p < 0.05$. Correlation coefficients over the conventional level for negligible (.30) are marked in bold

Why would deictic and temporal framing, in particular, be important to solve the EF problems presented by the WCST? Why would they be useful for rule flexibility? Let us look at a concrete example and see if we can describe the relations within a network of reasoning. Take the example statement from participant five (Table 6). This person deictically relates (i.e., from the perspective of “I”) possible present behavior, with previous behavior, via temporal framing. I-now is related temporally with I-then (literally “I-already,” in the example). The conclusion is to try the same behavior now as then. In the next paragraph, from the perspective of I-now, the participant relates previous behavior with a temporal frame, I-then (literally “I-the first time,” in the example). Using coordination and temporal framing, the participant relates previous behavior with a specific stimulus function (i.e., “what was correct-the first time”) and brings it back to guide behavior in the present. As is apparent in this example, temporal and deictic framing are not used exclusively. We have also coded several instances of coordination and distinction framing, and it is clear that these are also building blocks for rule flexibility. But our data suggest that deictic and temporal framing play the leading role whereas the other families are supporting actors.

Limitations

A basic premise for the present study was relying on the inference of function from the topography of language. This undoubtedly has its drawbacks, in terms of lack of experimental control. We agree with Atkins and Styles (2016), who used similar methodology in a study measuring rules in what people say, that this is a limitation.

It is clear that the present method of coding and analysis was unable to detect all principal families of framing. The family of opposition was seemingly absent from these participants’ reasoning. We do not conclude, based on this, that opposition framing as a verbal behavior was absent from participants’ actual pattern of thinking. We merely note that with the present method of analysis, relying completely on verbal statements as approximations of thinking, and in the context of the WCST, opposition framing was not visible.

The nature of this study was purposely exploratory. We wanted to explore which ways of thinking and which ways of categorizing might be useful, and our goal was also to describe the steps we took, even those that might have been less useful. One such example is that we zealously questioned participants and categorized their statements. In the end, however, we ended up only using category shift in our analyses. This decision was reached through discussion among all three authors. We considered this category most central in terms of the essence of EF. Of course, other choices could have been made in this respect.

The present study is an exercise relying heavily on interpretation. We realize that the lack of a formal procedure for interrater reliability checks in coding is a limitation. The method of RFT coding and analysis was constructed through back and forth discussion among all three authors. At an early stage, we experimented with independent coding of a fictional text, based only on our respective RFT knowledge (i.e., without using specific topographic cues). We found, however, that it was difficult to establish a reliable methodology. After further discussion, we started to rely increasingly on specific topographic cues, and reached consensus regarding the overarching method of coding, which is the one described in the methods section. However, while coding the texts, it was clear that even with a number of specified topographic cues, coding would also require idiographic functional analysis. Some specified topographic cues were subtracted because they were not functionally coherent. Also, it became clear that additional topographic cues would be added continuously because they appeared to some extent during coding itself. We therefore chose to let the first author conduct the majority of coding, and thereafter let author two scrutinize the texts for accuracy. During coding, the family of deictic framing in particular proved to be challenging in terms of deciding when it was present in participants’ statements about their thinking. This was confirmed by the relatively large number of additional instances of deictic framing that was added after author two had gone through the texts.

An additional limitation related to interrater reliability was the fact that the first author transcribed all sound recordings.

Table 6 Example reasoning from each participant

Participant	Correct new sorting rule in effect	Reasoning, verbatim
1	Shape	What do you think you can do now? It was color <u>just now</u> and <u>now</u> I could try if it is numbers, I mean the number of symbols, or perhaps symbols, if it is symbols <u>this time</u> .
2	Shape	What do you think you can do now? I could imagine that, if there has been a shift in the program, that <u>after</u> this amount of cards that I have placed that it has shifted to another combination. That it is no <u>longer</u> color.
3	Color	Why do you think that was incorrect? I'm thinking that I am trying, since I, since it was sort of like, it was <u>first</u> colors, and <u>then</u> symbols, and <u>then</u> numbers, that it like starts over, but perhaps it doesn't.
4	Shape	And why was that incorrect? I think that it goes for shape <u>again</u> . Ok. How come you think so? Because that was the <u>next</u> step that I took.
5	Color	What do you think you can do now? Same, same, yeah. I don't know what I am going to try <u>now</u> , actually. Something I have <u>already</u> tried. I am going to try the same color. Yes, and how come you choose that? Well, it might have been what I tried the <u>first</u> time. Or I didn't, but that was what was correct the <u>first</u> time.
6	Number	Why do you think that was incorrect? That's a good question. There is a one, perhaps there should, <u>now</u> I have placed them in color, I have placed them in triangles, and round dots. Should that one perhaps lay so that it is the same card <u>this time</u> , as I thought from the <u>beginning</u> ? If I try it like that <u>this time</u> . Correct.
7	Shape	Why do you think that was incorrect? <u>Now</u> I don't know if, <u>now</u> it feels like they are perhaps looking for some other strategy, that perhaps it shouldn't be sorted according to color any <u>longer</u> . They have changed. Right. And what do you think you can do? I'm thinking <u>again</u> that then I should perhaps go back and see <u>this time</u> if they want to sort it according to shape instead.
8	Number	What do you think you can do with the next one? The <u>next</u> one I'm thinking numbers. Ok. How come you think so? I mean that hasn't been <u>before</u> .
9	Color	And why do you think that was incorrect? <u>Now</u> I'm thinking that it switched back. So that it is possibly color. Yes, so I'm going to try to place it here.
10	Shape	Why do you think that was incorrect? <u>Now</u> it feels like it has switched patterns <u>again</u> and <u>now</u> perhaps I should have placed it according to how many symbols there are on the main cards. How come you think so? Because <u>now</u> it sort of feels like I have three different theories, and that is one of them.
11	Number	Why do you think that was incorrect? Well I guess it has switched systems <u>again</u> , perhaps. Yes. So what do you think you can do now? I guess I will try numbers <u>now</u> . How come you think so? Because I have <u>already</u> tried the other two.

Deictic cues are italicized, and temporal cues underlined. Test administrator's questions are in bold

Having an additional author transcribe the recordings and check for agreement would have added to the reliability of the data.

The small sample was a limitation. It was established arbitrarily, based on an estimate regarding how many participants would be required for a general pattern to be visible, in a qualitative sense. In future studies, based on the

nonsignificance of many of the correlations in the present study, a larger sample seems warranted.

To be clear, this work does not constitute a complete and ready system for how to code EF, but rather is meant as the first step of a research journey. The longer journey, we hope, will improve our potential for understanding central clinical issues and inspire ideas as to how we can intervene in a more efficient and theoretically sound way.

Future Directions

One obvious future line of study would be subjecting the proposed system of coding to formal interrater reliability testing. This would balance the rather heavy reliance on idiographic functional coding in the present work.

One reason it might be clinically important to clarify and concretize the concept of EF is that many clinical groups struggle with EF tests, such as the WCST. One such group is patients with schizophrenia (SZ). Studies (e.g., Harvey et al., 2005; Heaton et al., 1993; Kongs et al., 2000; Koren et al., 1998; Van der Does & Van den Bosch, 1992) consistently show inferior performance on the WCST in SZ patients compared to normal controls. Further, SZ patients' WCST performance seems to be stable despite repeated training (Harvey et al., 2005; Laws, 1999) whereas that of normal controls improve (Basso et al., 2001). There is also evidence for the correlation between WCST performance and work function (Greig et al., 2007; Lysaker et al., 2005) as well as psychosocial function (Kurtz & Wexler, 2006). Patients with ADHD and autism represent two other clinical groups shown to be prone to EF difficulties (Barkley & Murphy, 2011; Braden et al., 2017; Brown et al., 2009; Johnston et al., 2019). Clarifying in behavioral terms what constitutes EF could give some insight into behavioral deficits in those who do not do well on EF tests. Based on the results from the present study, a hypothesis might be that on average they do not use deictic and temporal framing during WCST performance to the extent that normal controls do.

Conclusion

Relational responding patterns during the WCST were dominated by coordination framing. However, during critical time points (i.e., at category shift when the need for EF is arguably most acute), the pattern changed. Coordination and spatial framing were used significantly less compared to during the whole test, whereas deictic and temporal framing were used significantly more. In terms of associations with outcome, there was a relatively clear pattern of positive correlations (small to large) between using deictic and temporal framing at category shift, and outcome variables of the WCST, whereas the correlations between coordination and spatial framing at category shift were negative. In conclusion then, we hypothesize that the concept of EF as measured by the WCST, in RFT terms, to some extent depends on deictic and temporal framing. Put in colloquial terms: to solve problems where EF is needed, people need to shift back and forth from a perspective of I-now, to I-then, I-before, etc. Although this might sound common-sense enough, we believe it might be a first step toward formulating EF deficits as a condition amenable to behavior change.

Availability of Data and Material The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability Not applicable.

Authors' Contributions All authors made substantial contributions to the design, analysis and interpretation of the present work. The first draft of the manuscript was written by Mårten J Tyrberg. All authors revised the work for important intellectual content. All authors approved the version to be published. All authors agree to be accountable for all aspects of the work.

Funding Open access funding provided by Uppsala University. This work was supported by the Centre for Clinical Research at Hospital of Västmanland, Västerås, Sweden (grant number LTV-906071).

Declaration

Conflicts of Interest The authors declare no conflicts of interest.

Informed consent Study procedures were approved by the regional Ethics Board at Uppsala University (Dnr 2017/130). All participants signed written consent forms at the time of inclusion. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atkins, P. W. B., & Styles, R. G. (2016). Measuring self and rules in what people say: exploring whether self-discrimination predicts long-term wellbeing. *Journal of Contextual Behavioral Science*, 5(2), 71–79. <https://doi.org/10.1016/j.jcbs.2016.05.001>.
- Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. Guilford Press.
- Barkley, R. A., & Murphy, K. R. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to performance on EF tests. *Journal of Psychopathology & Behavioral Assessment*, 33(2), 137–158. <https://doi.org/10.1007/s10862-011-9217-x>.
- Barnes-Holmes, D., O'Hora, D., Roche, B., Hayes, S. C., Bissett, R. T., & Lyddy, F. (2001). Understanding and verbal regulation. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 103–118). Kluwer Academic.
- Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the Wisconsin Card Sorting Test-64 Card version across 12 months. *The Clinical Neuropsychologist*, 15(4), 471–478. <https://doi.org/10.1076/clin.15.4.471.1883>.

- Belisle, J., Paliliunas, D., Dixon, M. R., & Tarbox, J. (2018). Feasibility of contextual behavioral speech analyses of US presidents: Inaugural addresses of Bill Clinton, George W. Bush, Barack Obama, and Donald Trump, 1993–2017. *Journal of Contextual Behavioral Science*, 10, 14–18. <https://doi.org/10.1016/j.jcbs.2018.07.002>.
- Braden, B. B., Smith, C. J., Thompson, A., Glaspy, T. K., Wood, E., Vatsa, D., Abbott, A. E., McGee, S. C., & Baxter, L. C. (2017). Executive function and functional and structural brain differences in middle-age adults with autism spectrum disorder. *Autism Research*, 10(12), 1945–1959. <https://doi.org/10.1002/aur.1842>.
- Brown, T. E., Reichel, P. C., & Quinlan, D. M. (2009). Executive function impairments in high IQ adults with ADHD. *Journal of Attention Disorders*, 13(2), 161–167. <https://doi.org/10.1177/1087054708326113>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *BRIEF: Behavior Rating Inventory of Executive Function: Professional manual*. Psychological Assessment Resources.
- Greig, T. C., Zito, W., Wexler, B. E., Fiszdon, J., & Bell, M. D. (2007). Improved cognitive function in schizophrenia after one year of cognitive training and vocational services. *Schizophrenia Research*, 96(1–3), 156–161. <https://doi.org/10.1016/j.schres.2007.07.003>.
- Harvey, P. D., Palmer, B. W., Heaton, R. K., Mohamed, S., Kennedy, J., & Brickman, A. (2005). Stability of cognitive performance in older patients with schizophrenia: An 8-week test-retest study. *American Journal of Psychiatry*, 162(1), 110–117.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Kluwer Academic.
- Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21–50). Kluwer Academic.
- Hayes, S. C., Gifford, E. V., & Ruckstuhl Jr., L. E. (1996). Relational frame theory and executive function. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory and executive function* (pp. 279–305). Paul H. Brookes.
- Hayes, S. C., Gifford, E. V., Townsend, R. C., & Barnes-Holmes, D. (2001). Thinking, problem-solving, and pragmatic verbal analysis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 87–102). Kluwer Academic.
- Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the “silent dog” method of analyzing the impact of self-generated rules. *Analysis of Verbal Behavior*, 15, 57–63.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised*. Psychological Assessment Resources.
- Hughes, S., & Barnes-Holmes, D. (2016a). Relational frame theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 129–178). Kluwer Academic.
- Hughes, S., & Barnes-Holmes, D. (2016b). Relational frame theory: Implications for the study of human language and cognition. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 179–226). Kluwer Academic.
- Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science*, 1(1–2), 17–38. <https://doi.org/10.1016/j.jcbs.2012.09.003>.
- Johnston, K., Murray, K., Spain, D., Walker, I., & Russell, A. (2019). Executive function: Cognition and behaviour in adults with autism spectrum disorders (ASD). *Journal of Autism & Developmental Disorders*, 49(10), 4181–4192. <https://doi.org/10.1007/s10803-019-04133-7>.
- Kibby, M. Y., Schmitter-Edgecombe, M., & Jong, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the California verbal learning test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology*, 13(6), 523–534.
- Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Wisconsin Card Sorting Test-64 card version professional manual*. Psychological Assessment Resources.
- Koren, D., Seidman, L. J., Harrison, R. H., Lyons, M. J., Kremen, W. S., Caplan, B., Goldstein, J. M., Faraone, S. V., & Tsuang, M. T. (1998). Factor structure of the Wisconsin Card Sorting Test: Dimensions of deficit in schizophrenia. *Neuropsychology*, 12(2), 289–302.
- Kurtz, M. M., & Wexler, B. E. (2006). Differences in performance and learning proficiency on the Wisconsin Card Sorting Test in schizophrenia: Do they reflect distinct neurocognitive subtypes with distinct functional profiles? *Schizophrenia Research*, 81(2–3), 167–171. <https://doi.org/10.1016/j.schres.2005.09.003>.
- Laws, K. R. (1999). A meta-analytic review of Wisconsin Card Sort studies in schizophrenia: General intellectual deficit in disguise. *Cognitive Neuropsychiatry*, 4(1), 1–35.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). Oxford University Press.
- Lysaker, P. H., Bryson, G. J., Davis, L. W., & Bell, M. D. (2005). Relationship of impaired processing speed and flexibility of abstract thought to improvements in work performance over time in schizophrenia. *Schizophrenia Research*, 75(2–3), 211–218. <https://doi.org/10.1016/j.schres.2004.09.014>.
- Mangeot, S., Armstrong, K., Colvin, A. N., Yeates, K. O., & Taylor, H. G. (2002). Long-term executive function deficits in children with traumatic brain injuries: Assessment using the Behavior Rating Inventory of Executive Functioning (BRIEF). *Child Neuropsychology*, 8, 271–284. <https://doi.org/10.1076/chin.8.4.271.13503>.
- Perrine, K. (1993). Differential aspects of conceptual processing in the Category Test and Wisconsin Card Sorting Test. *Journal of Clinical & Experimental Neuropsychology*, 15(4), 461–473. <https://doi.org/10.1080/01688639308402571>.
- Shute, G. E., & Huertas, V. (1990). Developmental variability in frontal lobe function. *Developmental Neuropsychology*, 6, 1–11. <https://doi.org/10.1080/87565649009540445>.
- Stewart, I., Barrett, K., McHugh, L., Barnes-Holmes, D., & O’Hora, D. (2013). Multiple contextual control over non-arbitrary relational responding and a preliminary model of pragmatic verbal analysis. *Journal of the Experimental Analysis of Behavior*, 100, 174–186.
- Stewart, I., Slattery, B., Chambers, M., & Dymond, S. (2018). An empirical investigation of part-whole hierarchical relations. *European Journal of Behavior Analysis*, 19(1), 105–124. <https://doi.org/10.1080/15021149.2017.1416525>.
- Van der Does, A. J. W., & Van den Bosch, R. J. (1992). What determines Wisconsin Card Sorting performance in schizophrenia. *Clinical Psychology Review*, 12, 567–583.
- Wicks, S., Hammar, J., Heilig, M., & Wisén, O. (2001). Factors affecting the short-term prognosis of alcohol dependent patients undergoing inpatient detoxification. *Substance Abuse*, 22(4), 235–245.
- Wilson, K. G. (2016). Contextual behavioral science: Holding terms lightly. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 62–80). Kluwer Academic.