Arbitrarily Applicable Relational Responding

Robert Johansson[1,2]

[1] Department of Computer and Information Science, Linköping University, Sweden
[2] Department of Psychology, Stockholm University, Sweden
robert.johansson@liu.se

Abstract. The purpose of this paper is to introduce how contemporary behavioral psychology approaches intelligence and higher-order cognitive tasks, as instances of so-called *arbitrarily applicable relational responding (AARR)*. We introduce the contemporary theory Relational Frame Theory (RFT), that suggests that key properties of AARR are *mutual entailment*, *combinatorial entailment*, and *transformation of stimulus function*. Furthermore, AARR are *contextually controlled* and developed through *multiple-exemplar training*. We explain these concepts and provide examples of how RFT uses this framework to explain complex cognitive tasks such as language, analogies, a sense of Self, and implicit cognition. Applications of RFT are surveyed. Finally, the relevance of RFT for the AGI audience is discussed.

Keywords: Relational Frame Theory · Behavioral psychology · Cognition · Intelligence · Language · Higher-order cognition · Self.

1 Introduction

In 1971, Murray Sidman was working with language comprehension with severely developmentally disabled individuals. Unexpectedly, he discovered that if subjects were successfully taught to match pictures and printed words to dictated words (AB and AC relations, respectively; Figure 1), and to name pictures (BD), they would without explicit training learn how to match printed words to pictures (BC), match pictures to printed words (CB) and to "read" (i.e., name words; CD). From a behavioral psychology point of view, this was very interesting, as it demonstrated a clear example of emitted behavior without a history of reinforcement.

2 Stimulus equivalence and derived stimulus relations

The above discovery has resulted in over 40 years of research in stimulus equivalence [10]. Stimulus equivalence is a behavioral phenomenon that (with one possible exception) seems to be limited to humans with verbal abilities. The possible exception is a california sea lion Rio, that seems to have demonstrated stimulus equivalence [9]. One way to study stimulus equivalence is with the help

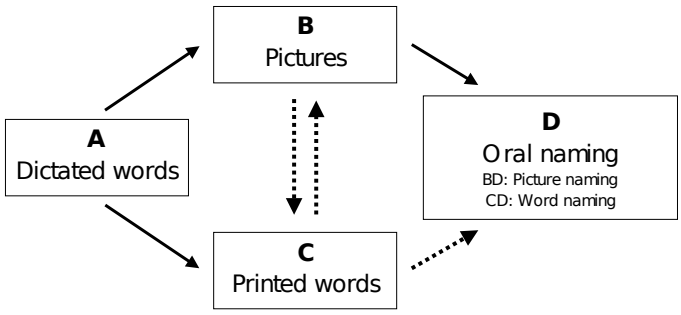

Fig. 1. The relations taught and tested by Sidman in 1971. Trained relations are depicted by solid arrows and derived relations with dashed arrows. Subjects were taught to match pictures and printed words to dictated words (AB and AC, respectively), and to name pictures (BD). Without explicit training they then could match printed words to pictures (BC), match pictures to printed words (CB) and to name words (CD).

of matching-to-sample experiments. In such experiments, participants are exposed to series of arbitrary stimuli (e.g., nonsense symbols) where the task is to match a certain symbol to a given sample stimuli. Such experiment is an example of *relational responding*. That is, the task for a participant is not to emit a response in relation to a certain stimulus. It is rather to respond to the relation between symbols.

A formal definition of stimulus equivalence follows. Assume three nonsense symbols, which we for simplicity will refer to as A, B and C (they might be nonsense words, pictures, or something else). Within a given experiment (like the matching- to-sample), participants are taught to select B rather than some other option in the presence of a sample A (i.e., the relation $A \rightarrow B$ will be established through training). In the same way C is trained as the correct response in the presence of B ($B \rightarrow C$). After these relations have been trained, without training in other relations, participants demonstrate an increased probability of selecting A from a set of options when B is presented as a sample ($B \rightarrow A$; symmetry), selecting C when A is displayed ($A \rightarrow C$; transitivity), selecting A when C is displayed ($C \rightarrow A$; equivalence), and also the trivial case of selecting A when A is displayed ($A \rightarrow A$; reflexivity).

Demonstrating symmetry and equivalence are examples of *derived relational responding*, as these stimulus relations are not directly taught but instead derived. Prior to the research by Sidman and colleagues [11] the emergence of these derived stimulus relations was not expected in similar experimental setups. As mentioned above, stimulus equivalence has been very difficult to demonstrate in nonhuman animals (except for the single sea lion). However, there exist research that has demonstrated symmetry in pigeons, monkeys and rats, but the results are somewhat inconclusive [7].

The stimulus equivalence phenomenon opened up for a new way of studying symbolic relations (i.e., how a word "represents" an object in language), and supported the idea that derived stimulus relations were an important component in language and cognition. Importantly though, the idea is not new. William James did already in 1890 regard the abstract concepts of sameness or equivalence as *"the very keel and backbone of our thinking"* [6, p. 459].

3 Arbitrarily applicable relational responding

In the late 1980's, the developers of Relational Frame Theory (RFT) [5] started to ask questions on what was beyond equivalence, for example: What kind of derived relational responding based on other relations than equivalence are human beings capable of? And if so, would such responding also be reflexive, symmetrical, and transitive? For example, consider a situation where someone is showed three identically sized coins, and being told that *"A is worth more than B, and B is worth more than C"*. Not only are the AB and BC relations specified, the BA, CB, AC, and CA relations will immediately be derived. Hence, a question such as *"Is C worth more than A?"* will be possible to answer (The answer would be *"No"*). Not only is this an example of responding to another relation than equivalence (a comparative relation), this is, according to RFT, an example of *arbitrarily applicable relational responding (AARR)*, as A, B and C are related along an arbitrary dimension of worth. In RFT terms, stimulus equivalence (as defined above) could be said to be a special case of AARR [15]. RFT has introduced more generic terms to describe features of derived relational responding, than the ones used to describe stimulus equivalence: *Mutual entailment*, *Combinatorial entailment*, and *Transformation of stimulus function*.

3.1 Mutual entailment

Like symmetry, *mutual entailment* refers to the fact that arbitrarily applicable relations are always bidirectional. If A is related to B, than a second relation BA is automatically entailed. The type of relation entailed depends on the relation between the two stimuli. For example, as illustrated above, if A is worth more than B, then the novel relation *"B is worth less than A"* is entailed. Another example would be, if A is the opposite to B, then B is also the opposite to A. In the latter case the same relation as the one trained would be entailed.

3.2 Combinatorial entailment

In line with transitivity, if A is taught to be related to B, and B to C, then a relation between A and C is *combinatorially entailed*. This was illustrated above where the *"A is worth more than C"* statement was derived. Once again, the type of relation entailed doesn't need to be the same as the one trained. For example, if someone is taught that *"A is the opposite to B"* and *"B is the opposite to C"*, then *"A is the same as C"* is combinatorially entailed.

3.3 Transformation of stimulus function

If A and B are taught to be related, and a response function (such as appetitive or aversive) is established for A, then the function of B will be *transformed* in accordance with the AB relation. For example, if someone fears dogs and learns that the word *"hund"* means *"dog"*, then the aversive stimulus function of *"dog"* is predicted to transfer through the sameness relation of *"means"*. Another example of non-equivalence follows. If someone learns that two nonsense stimuli are related $A < B$, and A then is paired with a mild electric shock, then the stimulus function of A will be transformed from neutral (as for a nonsense symbol) to aversive. Importantly though, B is predicted to be transformed to even more aversive. Similar effects have been demonstrated experimentally, using skin conductance equipment [3].

3.4 Contextual control over AARR

Consider the example above, with a person seeing three identically sized coins, learning that *"A is worth more than B, and B is worth more than C"*. Imagine that the person instead learned that *"A is worth the same as B, and B is worth the same as C"*. In the two situations, two different forms of AARR would be triggered, for example as part of a decision-making scenario involving money. More specifically, the *"more than"* and *"same as"* are two different forms of *contextual cues*. This highlights the contextual nature of AARR.

3.5 Multiple-exemplar training

How is arbitrarily applicable relational responding developed during lifetime? RFT assumes this is due to a history of *multiple exemplar training*. Imagine for example a small child who hasn't learned to apply the concept of comparison. Through interaction with the environment, the child might hear that *"the horse is larger than the duck, and the duck is smaller than the horse"*, and *"the man is longer than the child, and the child is shorter than the man"*, etc. RFT assumes that these multiple examples over time leads to the applicable abstract pattern of comparison that fulfills the properties of relational frames mentioned above.

3.6 AARR and relational frames

In summary, *arbitrarily applicable relational responding (AARR)* is defined as *abstract response patterns*, that have the properties of *mutual entailment*, *combinatorial entailment* and *transformation of stimulus functions*, that are controlled by *contextual cues* and learned through a history of *multiple exemplar training*.

Specific instances of AARR (for example sameness and comparison), are referred to as different types of *relational frames*. The term is based on a metaphor of a picture frame. Just as a picture frame can hold many pictures, a response frame can include many different features while still being a specific instance of an overall pattern.

4 Families of Relational Frames

In this section, we will elaborate further on RFT by describing the most common types of relational frames. Importantly, RFT is not limited to these frames. These are overall patterns that have been found useful to distinguish from one another.

Coordination A frame of *coordination* is essentially a relation of sameness. If someone is taught that *"A is the same as B"*, and *"B is the same as C"*, then the BA, CB, AC, and CA relations of sameness will be entailed. This is essentially the same as stimulus equivalence. Furthermore, if the person is taught *"C tastes disgusting"*, then the aversive stimulus function of C will transfer to A and B, both being about equally aversive. RFT research has suggested that sameness is the earliest relational frame to develop, and arguably the most fundamental. This seems related to the fact that children early in their development tend to learn that words "refers" to things and events, that is, being the "same as".

Opposition A more complex relational frame is that of *opposition*, that is relating stimuli in the presence of cues such as *"is opposite of"*. For example, a statement such as *"If Aaron (who is very tall) is opposite to Bill, and Bill is opposite to Charlie, then what is Charlie like?"* involves this frame. The statement needs to involve explicit or implicit information on which dimension along with the stimuli may be differentiated. These dimensions could be physical such as size, temperature, and brightness, but also arbitrary dimensions, as for example in the following statements: *"easy is the opposite of hard"*, *"valuable is the opposite of worthless"*, and *"A is opposite to B, and B is opposite to C. A is funny. Is C funny? Is B?"*.

Distinction The frames of *distinction* are controlled by cues such as *"is different from"* and *"is not the same as"*. For example, if A is taught to be of different color from B, then it is entailed that B is of different color from A. However, the frame of distinction doesn't have the same specificity as the previously described frames when the relational networks grow, as shown in the following example: *"A has a different color from B, and C has a different color from B. A is green. Is B green? (No) Is C green? (Don't know)"*.

Comparison *Comparative frames* involve responding to stimuli or events in terms of a quantitative or qualitative relation along some specified dimension. For example, *"If a dime is worth more than a nickel, and a nickel is worth more than a penny, is a dime worth more or less than a penny?"* is a statement which would require a person to derive the comparative relation between a dime and a penny. More specifically, the cue *"is worth more"* signals that the frame of comparison could be applied. Other examples of cues that control this kind of relating are *"heavier/lighter"*, *"better/worse"*, and *"larger/smaller"*.

Hierarchy These are frames involving membership, attributes, class containment, etc. For example, *"If an object A is a type of object B and an object B is a type of object C, then is an object A a type of object C?"*. Also consider this example: *"If coffee is a type of drink, and tea is a type of drink, is then coffee a type of tea?"* In that example, the relationship isn't specified.

Temporal frames Responding to events in terms of temporal displacement from other events represents an example of responding in accordance with temporal relations, such as *"before/after"*. For example, *"If Tuesday comes before Thursday, and Thursday comes before Friday, does Friday come before or after Tuesday?"*

Spatial frames These frames involve relating along a spatial dimension, and may be triggered by cues such as *"above/below"*, *"left/right of"*, *"here/there"*, *"front/back"*, etc. For example, given that *"If A is above B, and B is above C"*, a person will derive that *"C is below A"*, *"A is above C"*, *"B is above A"*, and *"C is above B"*.

Deictic frames Finally, *deictic* frames are those that specify a relation between stimuli from the perspective of the speaker. RFT suggests that these deictic frames are a combination of three types of relations: *spatial* (*"here/there"*), *temporal* (*"now/then"*), and *interpersonal* (*"me/you"*). An example of a statement involving deictic framing is *"If I am here and you are there, and if I were you and you were me, where would you be? Where would I be?"*. Another example is *"If I feel sad and you feel happy, and if I were you and you were me, how would you feel? How would I feel?"*. The latter could be said to be an example of how something such as empathy could be analyzed through RFT.

5 Cognition and intelligence from an RFT perspective

From an RFT perspective, *cognition* is not a mental event that mediates between environment and behavior. It is rather a behavioral event (AARR), and hence, it can be studied and understood within a behavioral psychology framework, using experiments such as the matching-to-sample task described above. Another way to put it: arbitrarily applicable relational responses are what "minds" are full of, and when we speak of "cognitive" phenomena (such as thinking, planning, remembering, decision making) we are referring to complex instances of relational framing that are more or less evident under different environmental conditions [15].

Regarding *intelligence*, the core idea from RFT is that AARR represents the basic functional "building block" of cognitive and linguistic skills, such as deductive and inductive reasoning, communication, etc., all of which underpin intelligent behavior. In essence, intelligent acts involve the ability to elaborate networks of derived stimulus relations fluently and flexibly, to transform stimulus

functions through entire networks, and to bring relational responses under increasingly subtle forms of contextual control, by abstracting relevant contextual features with high precision.

6 Higher-order cognitive tasks and complex relational responding

In this section, we will provide some examples on how RFT approaches various complex cognitive skills given the framework introduced above.

6.1 Language

RFT approaches language as verbal behavior [15]. A person learns "how to language" by learning how to respond relationally to stimuli and events. Hence, verbal behavior and language from an RFT perspective is really about the act of "framing events relationally". Stimuli such as words (spoken or written) or pictures become "verbal stimuli" when they participate in relational networks with contextual cues. It is this process that enables "meaning" to something as the stimuli acquire various stimulus functions. Someone speaks "with meaning" when they frame events relationally and produce sequences of verbal stimuli as a result. Someone else will "listen with understanding" whenever they respond as a result of framing events relationally. In essence, understanding something is not an outcome of an "inner/mediating" mental event, but is rather a type of contextually controlled behavior.

6.2 Analogies

All of the examples above have focused on how stimuli or events can be related. However, sets of relations can also be related. Relating relations is, from an RFT perspective, the basis of how analogies are developed and used [12]. For example, a quite simple analogy might be *"Apples are to oranges as dogs are to sheep"*. This can be described as an equivalence relation between equivalence relations. More specifically, apples and oranges participate in a relation of equivalence (fruits), while dogs and sheep also participate in a relation of equivalence (animals). An example of analogical reasoning given this is deriving these two equivalence relations and the derivation of another equivalence relation between the relations. In other words, apples are equivalent to oranges in the same way that dogs are equivalent to sheep, because they are members of the same respective class). A further example could be someone who already knows about the solar system, and is learning physics. The statement *"An electron is to the nucleus as a planet is to the sun"* involves an equivalence relation between spatial relations. Given that the person knows this relation between planets and the sun, he/she could then derive a new spatial relation between electrons and nuclei.

6.3 The Self and perspective taking

As described previously, deictic frames involve temporal (*"now/then"*), spatial (*"here/there"*), as well as interpersonal relations (*"Me/You"*). While coordination, distinction and comparative relations (see above) develop based on what people learn about stimuli that are physically similar, dissimilar, or quantitatively different along some dimension, deictic frames are typically not. Instead, they develop based on the invariance of the speaker's perspective throughout time and location. A child might learn this by being asked questions such as *"Who are you?"*, *"What are you doing here?"*, *"What will you do there?"*, and *"What will I do tomorrow?"*, with many variations, in several different contexts. By taking part of a constant relating of *"Me/You"*, *"here/there"*, *"now/then"*, a child learns about itself, as something being different from others, and being *"here and now"* as compared to *"there and then"*. Hence, in line with how RFT provides an understanding of "languaging" as framing events relationally, "selfing" is approached similarly. Understanding how it is possible to take someone else's perspective also follows naturally from this analysis [8].

6.4 Implicit cognition

To account for both "thinking fast and slow", RFT introduces dimensions to AARR such as *levels of complexity*, *derivation*, and *coherence* [1]. Complexity refers to the number of stimuli or events involved, with for example a mutually entailed response being "less complex" than a relating of relations. Levels of derivation is a continuum from a relating with very few new derivations on one end, and a response involving a large amount of new derivations on the other end. A response that is low is coherence is very little in agreement with a larger relational network that the response is taking place in. On the contrary, a response with high overlap with previous experience, is said to be high in coherence.

In an experimental task that studies implicit cognition from an RFT perspective, there is an assumption that responses that requires low levels of complexity and derivation, and being high in coherence, will be very quickly emitted. However, responses that require a high level of complexity and derivation, and/or being low in coherence, are assumed to be slower emitted, and therefore lead to longer response times.

7 Applications of RFT

Below, we will provide examples of how RFT can be used in applied settings, outside experimental psychology.

7.1 Education

It follows naturally from the above description, how teaching based on sameness, opposition, comparison, etc, could be conducted. The importance of multiple-exemplar training is highlighted by RFT. Furthermore, RFT provides the tools

on how various new relational networks could be established, set in relation to existing networks, affected by transformation of function, etc. RFT provides the details on how analogies could be used as part of education, and provides an account of experiential learning through transformation of stimulus function. Furthermore, skills training involving perspective taking, such as training in empathy, could be understood through the lens of RFT [13].

7.2 Clinical applications

Human suffering seems to be very much related to our capacity for language [2]. Statements such as *"Deep down I'm a bad person"* or *"I am not worthy of love"* or *"Everybody else is much better than me in most aspects"* are common in depression and related problem presentations. For anxiety disorders, simple statements such as *"Spider"* can trigger a whole host of physiological reactions. Similarly, when other terms are taught to be in equivalence with *"Spider"*, then these terms are predicted to trigger similar reactions. Furthermore, RFT provides an understanding of how other stimuli and events can become closely related to spiders, which could result in that the fear will generalize to other similarly looking things [4]. In summary, RFT provides accounts of a whole range of clinical phenomena, and provides tools on how to resolve these issues.

7.3 Prejudices

Today's society undoubtedly face massive problems related to hate, discrimination and violence due to prejudice. From an RFT perspective, prejudice could be defined as objectification and dehumanization of individuals because of their participation in verbal evaluative categories [14]. A major challenge seems to be due to the fact that prejudice and related processes seem to stem from the same source as our most successful problem-solving processes. RFT might be able to provide the means to deal with this verbal entanglement.

8 How could RFT be relevant for AGI researchers?

One could argue that RFT is essentially a behavioral psychology approach to general intelligence. While the AGI field has benefited from theories from diverse fields such as computer science and neuroscience, we believe that behavioral psychology also has something to offer. In the complex task of building thinking machines, clear definitions of cognitive phenomena are likely to be very helpful. RFT suggests that AARR is a necessity for intelligence and higher-order cognitive tasks. Possibly, RFT could provide a roadmap based on a science of derived stimulus relations, starting with symmetry and stimulus equivalence, going up to relations between relational networks, with models of language development, higher-order cognitive tasks, and the Self, with potential applications within diverse fields such as education, psychological treatments, and prejudices. We hope by writing this text that the AGI field finds such roadmap potentially helpful.

References

1. Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., McEnteggart, C.: From the irap and rec model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. Journal of contextual behavioral science **6**(4), 434–445 (2017)
2. Barnes-Holmes, Y., Barnes-Holmes, D., McHugh, L., Hayes, S.C.: Relational frame theory: Some implications for understanding and treating human psychopathology. International Journal of Psychology and Psychological Therapy **4**, 355–375 (2004)
3. Dougher, M.J., Hamilton, D.A., Fink, B.C., Harrington, J.: Transformation of the discriminative and eliciting functions of generalized relational stimuli. Journal of the Experimental Analysis of Behavior **88**(2), 179–197 (2007)
4. Dymond, S., Schlund, M.W., Roche, B., Whelan, R.: The spread of fear: Symbolic generalization mediates graded threat-avoidance in specific phobia. Quarterly Journal of Experimental Psychology **67**(2), 247–259 (2014)
5. Hayes, S.C., Barnes-Holmes, D., Roche, B.: Relational frame theory : a post-Skinnerian account of human language and cognition. Kluwer Academic/Plenum Publishers, New York (2001)
6. James, W.: The principles of psychology, Vol I. Henry Holt and Co, Inc, New York, NY (1890)
7. Lionello-DeNolf, K.M.: The search for symmetry: 25 years in review. Learning & behavior **37**(2), 188–203 (2009)
8. McHugh, L., Stewart, I.: The self and perspective taking: Contributions and applications from modern behavioral science. New Harbinger Publications (2012)
9. Schusterman, R.J., Kastak, D.: A california sea lion (zalophus californianus) is capable of forming equivalence relations. The Psychological Record **43**(4), 823–839 (1993)
10. Sidman, M.: Equivalence relations and behavior: An introductory tutorial. The Analysis of verbal behavior **25**(1), 5–17 (2009)
11. Sidman, M., Tailby, W.: Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. Journal of the Experimental Analysis of behavior **37**(1), 5–22 (1982)
12. Stewart, I., Barnes-Holmes, D.: Relational frame theory and analogical reasoning: Empirical investigations. International Journal of Psychology and Psychological Therapy **4**(2), 241–262 (2004)
13. Vilardaga, R.: A relational frame theory account of empathy. International Journal of Behavioral Consultation and Therapy **5**(2), 178 (2009)
14. Weinstein, J.H., Wilson, K.G., Drake, C.E., Kellum, K.K.: A relational frame theory contribution to social categorization. Behavior and Social Issues **17**(1), 40–65 (2008)
15. Zettle, R.D., Hayes, S.C., Barnes-Holmes, D., Biglan, A.: The Wiley handbook of contextual behavioral science. Wiley Online Library (2016)