13e CONGRÈS INTERNATIONAL DE CYBERNÉTIQUE

Namur (Belgique), 24-28 août 1992

13th INTERNATIONAL CONGRESS ON CYBERNETICS

Namur (Belgium), 24th-28th August 1992

ACTES

PROCEEDINGS

Namur

ASSOCIATION INTERNATIONALE DE CYBERNÉTIQUE

1992

OVERCOMING SUBJECTIVE BARRIERS IN LEARNING NORMS BY FEEDBACK

By G. PEETERS

1. INTRODUCTION

Action can be conceived of as guided by subjective norms within the actor. Success and failure then can be defined as feedback informing the actor about the (dis)agreement of the subjective with an objective norm. For instance, the embarrassing experience that the long-haired person I called "miss" turns out to be a boy, functions as right-wrong feedback informing me about the inadequacy of hair-length as a subjective norm for calling people "miss" because lacks agreement with the objective norm "gender".

Some data are presented showing how subjective norms are adjusted -- or fail to be adjusted -- by right-wrong feedback. Subjects (psychology graduate students) were confronted with 8 vignettes with on each vignette an item telling how 2 hypothetical people (A and B) related to each other and to themselves. The items, labelled a,d,.. p, are displayed in the INPUT part of Tab.1 where + and - represent positive and negative relations. Thus item (vignette) f tells that A relates positively to B, B relates negatively to A, A relates positively to the self and B relates negatively to the self. In the experiments + and - relations were encoded respectively as "liking" and "disliking"; e.g., vignette f offered the following text:

"A likes himself; A likes B; B dislikes himself; B dislikes A"

The vignettes were presented one by one to the subject who had to respond whether A and B seemed "alike" or "different" (OUTPUT). Each response was followed by "right-wrong" feedback consistent with one of the three NORMS in the bottom part of Tab.1. The procedure was repeated until the subject had learned correct responses to all items. Then the subject had to explain the rule (norm) he or she thought the feedback was based on.

Consistent with a previously observed *Self-Other-Anchoring Bias* in relational information processing (1), subjects were found to start with NORM 1 (e.g.: A and B alike if both like the self). Thus only feedbacks consistent with norms 2a and 2b were used. Those norms were called *"3P-Anchored"* in that the (dis)similarities between A and B were not defined by how they related to self and other, but by how they related to specific individuals conceivable in terms of the third pronominal person such as *"Him or her called X"*. Further, the relations being vectors, (dis)similarities between A and B could be defined either at *origins* or at *terminals*. The origin-terminal distinction did not matter for SO-anchored (dis)similarities but for 3P-anchored ones we had to distinguish between terminal- and origin-associated norms 2a and 2b. The (dis)similarities associated with the norms are presented as OUTPUTS in Tab.1 (rows 5-7) where + and - now represent respectively *alike* and *different*. Thus for item f, - (row 5) means that A and B are different from each other following NORM 1; + (row 6) means that they are alike following norm 2a, and - (row 7) that they differ from each other following norm 2b.

G. PEETERS, Research Associate N.F.W.O., K.U.Leuven, Laboratoriumm voor Experimentele Sociale Psychologie, Tiensestraat, 102, B-3000 Leuven, Belgium

Table 1. *Input, Output and Norms concerning the items a, d,.. p.*

ITEM LABELS:		a	d	f	g	j	k	m	p	
INPUT:	Relation AB:	+	+	+	+	-	-	-	-	row 1
	Relation BA:	+	+	-	-	+	+	-	-	row 2
	Relation AA:	+	-	+	-	+	-	+	-	row 3
	Relation BB:	+	-	-	+	-	+	+	-	row 4
OUTPUT 1 (following NORM 1):		+	+	-	-	-	-	+	+	row 5
OUTPUT 2a (following NORM 2a):		+	-	+	-	-	+	-	+	row 6
OUTPUT 2b (following NORM 2b):		+	-	-	+	+	-	-	+	row 7

NORM 1 *(SO-Anchored)*: A and B alike when they relate in the same way to the self and/or in the same way to the other.

NORM 2a *(3P-Anchored and Terminal-Associated)*: A and B alike when they are related to in the same way by the same individual(s) X (note that X can be either A or B or both A and B).

NORM 2b *(3P-Anchored and Origin-Associated)*: A and B alike when they relate in the same way to the same individual(s) X (note that X can be either A or B or both A and B).

Before turning to human intelligence, we consider briefly how a machine intelligence might be programmed for the task. First, a PROGRAM 1 can be conceived involving 2 stages. In *Stage 1* the correct responses are inferred from the feedback and stored in memory. In *Stage 2* the stored responses are used to select a fitting norm. Second, according to 'PROGRAM 2', the machine starts with selecting a norm and uses it for generating responses. If responses are disconfirmed by feedback, another norm is tried out and so on until all of the generated responses are confirmed by the feedback. These machine models serve as straw-men to be confronted with human intelligence. For instance, our finding that subjects tended to start responding following NORM 1 and changed under the influence of the feedback is consistent with PROGRAM 2. However, inconsistent with PROGRAM 2, the subjects often failed to discover the right norm. Nevertheless they were able to acquire the correct responses, which seems consistent with *Stage 1* of PROGRAM 1 while *Stage 2* was not achieved. The lack of agreement between machine and human intelligence may be due to differences in the availability of the various norms that are to be selected. For the machine intelligence those norms may be readily available by having Tab.1 included in the programs. As to human intelligence, one may consider that the given norms look that self-evident that they must be readily available for the psychology graduate student. However this seems not to be the case. The explanations subjects gave about the rule they thought to underlie correct responses, showed that the availability of norms was inhibited by subjective barriers. Below we present some still unpublished data from psychological experiments (1) illustrating how subjective barriers may obstruct learning.

2. EXPERIMENTS

EXPERIMENT 1. 4 Subjects were instructed that (dis)similarities between A and B concerned A and B's *tastes and preferences.* Then they got feedback consistent with NORM 2b *(3P-anchored/Origin-associated).* 3 Of them achieved correct responding and formulated explanations in agreement with the norm, telling, for instance, that *A and B had the same tastes and preferences in*

that they adopted the same attitude towards A. It is not clear whether the subjects followed PROGRAM 1 or 2 but it is evident that NORM 2b was readily available. The 4th subject, however, failed to learn the correct responses. Afterwards he declared that in the beginning he tried to find a rule but that he did not see any system in the feedback and ended with random responding. Apparently the subject proceeded by PROGRAM 2, but NORM 2b was not available to him.

EXPERIMENT 2. Another 4 subjects were instructed that (dis)similarities concerned A and B's *tastes and preferences* but got feedback consistent with NORM 2a *(3P-anchored/Terminal-associated).* Subjects 1, 2, and 3 achieved correct responding. However, Subj.1 could not give a reasonable explanation and might have achieved the task by rote learning. Subj.2 declared that *similarity was assigned to A and B if the relation to the other was similar to the relation with the self.* Despite the use of the terms *self* and *other,* this message did not reflect the rationale of NORM 1 from which the subjects tended to proceed when starting the task. One could even argue that the message reflected the correct norm in that it implied that A relating in the same way to the self as to the other implies that A and B are related to in the same ways by the same individual. However, the subject continued her explanation with the specification that A and B were, for instance, similar *when they related in the same way to A.* This was clearly a 3P-rationale, indeed, but it reflected NORM 2b rather than the correct NORM 2a. Subj.3 declared that A and B were alike *if they adopted the same attitude towards A,* which once more reflected NORM 2b. Subj.4, who failed to achieve correct responding, was found to produce responses consistent with NORM 2b from the 2nd repetition of the 8-item series on and he persisted during 7 repetitions until the session was broken off because subject and experimenter got exhausted. He declared, consistent with NORM 2b, that he had considered A and B as similar *when they adopted the same attitude towards A* and that this rationale appeared so *natural* that it was *impossible to find any other rule.* Conclusions are:

1.- Consistent with PROGRAM 2, the subjects learned to drop NORM 1, but they failed to formulate NORM 2a.

2.- Instead of NORM 2a, Subjects 2, 3, 4 derived the more available but wrong NORM 2b which was 3P-anchored like the right NORM 2a. Apparently it was the terminal-bound character of NORM 2a that formed an insuperable barrier.

3.- Although they derived an inappropriate norm, Subjects 2 and 3 managed to learn the correct responses consistent with *Stage 1* of PROGRAM 1. This amazing discrepancy between the learning of the responses and the learning of the norm suggests that in human intelligence parallel learning programs can go on that can even lead towards logically incompatible outcomes.

EXPERIMENT 3. One possible explanation for the strong origin-associated outputs observed in experiment 2 may reside in the origin-associated nature of attitudinal categories such as "tastes and preferences". Indeed, it seems evident to interpret "A liking B" directly as a preference of A rather than as an indirect consequence of some hidden preference of B. However, if the (dis)similarities to be assigned to A and B would concern their personalities, then the barrier against NORM 2a may be abolished. For instance, given item f one can easily imagine A and B as alike in personality (e.g.: two cholerics) but with different preferences (A liking cholerics while B disliking them). Hence 4 subjects were instructed that the (dis)similarities between A and B concerned their personalities and feedback was given consistent with NORM 2a. 2 Of them, one of whom had successfully learned the correct responses, could not offer any reasonable explanation. The third, who also was successful in learning the correct responses, declared that *when A and B were consistent in*

their relations with oneself and the other, then they showed the same degree of constancy and thus were alike in character; if both were not consistent in their relations, then they were different in character.

The latter subject seemed to advance a new norm that we may call *UNIFORMITY NORM*. It states that A and B are alike if they show the same degree of uniformity in the relations emanating from them. Relations are uniform when they have the same sign. Thus if A likes both the self and B, and B dislikes both the self and B, then A and B are alike in that both are uniform in their relations. The subject called this uniformity "consistency" and he associated it with the personality trait "constancy". An important point is that this norm does not lead to the correct response pattern achieved by the subject, but to "alike" for all items, which could be represented in Tab.1 by an output-row with only + signs. However, the subject did not draw this consequence. Instead he interpreted only "similarity in high constancy" as "similarity" while "similarity in low constancy" was called "dissimilarity", which was very illogical, indeed, but led to correct responses. Thus the subject achieved success by applying a wrong norm in a wrong (illogical) way.

The 4th subject showed more logical consistency but did not stick to one single norm. Actually the following norms were used:
1.- the above-mentioned *uniformity norm* to items a, f, k, and p *(A and B similar in that both show consistency over their attitudes to self and other);*
2.- NORM 1 to items g and j *(A and B different in that they adopt different attitudes towards the self);*
3.- a *"Liking-Similarity" norm* to items d and m *(A and B alike if they like and different if they dislike each other).* This norm has been shown to be unrelated to the other norms mentioned so far (1).

The main outcome was that focussing on personality, the quite self-evident terminal-bound responses were not obtained. The subjects even failed to formulate any 3P-anchored norm. Instead they advanced new norms that were complex and even logically inconsistent.

EXPERIMENT 4. The "personality" instruction of experiment 3 was used with another 4 subjects but this time the feedback was consistent with the origin-bound NORM 2b. All subjects achieved correct responding but 3 of them could not provide any reasonable explanation. The 4th, however, reported the correct norm telling that *A and B had the same character when they related in the same way to A.* He added that *he could comprehend the given rule but would never have applied it spontaneously.*

The outcomes of experiments 3 and 4 may surprise but they are in full agreement with other studies confirming that the implicit personality concept underlying non-professional personality judgments is not only predominantly origin-bound but also strongly SO-anchored (1). Apparently the intrinsic SO-anchored nature of the implicit personality concept constitutes an insuperable barrier for discovering 3P-anchored norms on the mere basis of right-wrong feedback. Apparently this barrier belonged to a level of implicit rather than of explicit knowledge. Indeed, when the experimenter made the norm explicit, by just telling it, subjects had no problem to understand it and were surprised that they had been unable to discovered it themselves. Effects of explicit revelation of the norm were further explored in experiment 5.

EXPERIMENT 5. In an early pilot experiment 16 instead of 8 items had been used. Neither the correct responses nor the correct norms had been learned by any subject. However, at the end of the session the experimenter had explained the correct norms telling, for instance, that A and B were alike when they both liked B (NORM 2b) or when both were liked by B (NORM 2a).

One year after, the above experiment 4 was repeated with 3 subjects who in the pilot experiment had got feedback consistent with NORM 2b. Only 1 subject remembered the norm and applied it correctly from the first trial on. Subj. 2 remembered only that he had to use various combinations and that only part of the information included in an item was needed. Subj. 3 remembered only that she had tried to compare the self-images of A and B and that this had been a bad strategy. Both subjects 2 and 3 learned now the correct responses although not faster than in the above experiment 4. However, they reported the norm correctly. The first said that A and B were alike *when the attitude of A towards A and the attitude of B towards A were in the same direction.* The second said: *"If A and B relate in the same way to B, for instance, then this means that A perceives something in B that B perceives also in himself".*

Experiment 3 was repeated with 3 subjects who 1 year before had got feedback consistent with norm 2a. They also were able to learn the correct responses and did it not faster than the subjects in the above experiment 3. Subject 1 remembered from the year before only that no more than 2 out of the 4 statements in each item were to be used but she used nevertheless the 4 statements when she explained correctly that A and B were alike *when A said the same about himself as about B and B said the same about himself as about A.* Subject 2 remembered only that he had found the feedback strange and had not changed his initial view. Concerning the present session, he declared to have applied different rationales to different items among which there were glimpses of the UNIFORMITY NORM (see higher exp. 3), of NORM 1, and, last but not least, of NORM 2a on which the feedback was based. For instance, about item d he declared: *"A is negative with respect to A but nevertheless likes B; thus there must be a difference somewhere between them."* The last subject remembered only that she had not to make it herself too difficult. Her explanations showed glimpses of the NORM 1 and the UNIFORMITY NORM but not at all of the relevant NORM 2A. Conclusions are:

1.- Subjects to whom the norm has been revealed in the past do not perform better in response-learning except if they remember the norm explicitly.

2.- Nevertheless, even if they do not remember the norm, they have become more able to overcome the barriers set by the SO-anchored and origin-bound nature of the implicit personality concept.

3.- NORM 2b being more accurately inferred than NORM 2a confirms the origin-bound nature of the implicit personality concept.

3. CONCLUDING REMARKS

1.- Consistent with *Stage 1* of the above PROGRAM 1, correct responses can be learned without acquiring explicit understanding of an appropriate norm.

2.- In order to achieve *Stage 2* of PROGRAM 1, as well as in order to learn by PROGRAM 2, the availability of appropriate norms is required. This availability is detracted from by subjective barriers that belong to implicit knowledge about the nature of the responses (e.g.: a barrier against 3P-anchored norms built into the implicit personality concept)

3.- The barriers do not prevent learning correct responses following PROGRAM 2 in that subjects may bypass barriers by advancing alternative norms that may be unduly complex and even illogical but effective in generating right responses.

4.- When the right norm has been made explicit in the past but is no longer remembered, then response-learning by right-wrong feedback seems not improved. However the barriers that inhibit the discovery of the norm are weakened.

---

(1) Peeters, G. (1991). Relational information processing and the implicit personality concept. *European Bulletin of Cognitive Psychology, 11,* 259-278.