Psychology Science, Volume 49, 2007 (3), p. 255-270

Towards the development of behavioural measures of impulsivity: Influencing subjects' response style in a multiple-choice visual discrimination task

MICHAELA M. WAGNER-MENGHIN[1] & NICOLE K. DIETRICH

Abstract

In an attempt to reduce guessing during multiple-choice testing, the effects of different payoff matrices were observed in this research. It was hypothesized that a penalty for incorrect answers and the existence of the alternative "no decision" have a significant effect on reducing guessing by subjects. Furthermore, this effect should become stronger the higher the risk of an incorrect response and therefore the higher the risk of losing points is. An experimental design with four different payoff matrices and a repeated measurement design with cross over (for the payoff for "no decision") was realized. A visual discrimination task including three steps (one power condition, two speed conditions), with each item containing three choices plus the "no decision" option, was used. The results showed that the percentage of correct answers is influenced by the different payoff for the choice of "no decision". Only in groups where "no decision" paid off with 0, was it chosen more often, however, the percentage for correct answers is lower in these groups. The results also indicate that subjects are able to adjust their behaviour according to the different payoff matrices.

Key words: Impulsivity, payoff, behavioural measure

[1] Mag. Dr. Michaela Wagner-Menghin, Institut für Entwicklungspsychologie und Psychologische Diagnostik, Zimmer: O1.40, Liebiggasse 5, 1010 Wien, Austria; mail: michaela.wagner@univie.ac.at

Answering items randomly, preferring the first possibility over the others or guessing when in doubt are only some of the examples of several response styles we encounter in psychological assessment. Usually response styles are regarded as a problem in questionnaire based personality assessment as they distort the result of the questionnaire. However, the response style gambling vs. cautiousness (Cohen & Swerdlik, 1999, p.389) is also observed with ability tasks. This response style is a problem especially with multiple choice items and might also distort results. Especially in the case of speed and power tests guessing, when running out of time, might lead to a score disproportionate to the latent ability. The careful construction of distractors, offering "*no decision*" or "*I do not know the answer*" as an option (e.g. Gittler, 1990) and *losing points credits* when answering incorrectly (e.g. Amthauer, 1972) are just some of the possibilities mentioned in order to reduce the distorting effect of guessing. However the first possibility is said to be ignored by the testee, because nothing can be gained by admitting that one does not know the answer and the second is perceived as unfair, especially when the test administration requires the input of an answer for every item as it is the case with some computerized time limited tasks.

But maybe the key to success in dealing with the gambling vs. cautiousness problem lies not solely in trying to reduce it? The way an individual approaches a task and deals with uncertainty might provide valuable information about personality traits, especially for those traits that relate to behaviour in achievement situations e.g. endurance, aspiration level, impulsivity or conscientiousness. There may be potential in using information about a person's response style with achievement tasks as a variable in behaviour-based personality assessment.

The idea of behaviour-based personality assessment was introduced by Cattell's objective personality tests (OPT, Cattell, 1958; Cattell & Warburton, 1967) which rely on assessing a personality construct by observing the person's behaviour when working on a (performance or an ability) task, rather than by using self reports about behaviour. This approach has recently been resumed by Ortner, Proyer and Kubinger (2006) and Wagner-Menghin (2006b). This group deals with the development of tasks and the construction of measures for the assessment of achievement related traits, like learning styles, stress resistance, achievement motivation and impulsivity. Kubinger (2006), in particular, emphasizes the potential of experimental manipulation as an assessment principle, hence he refers to the procedures following this principle as experiment-based behavioural tasks (EBT). In referring to measures derived from these tasks we use the term "behavioural measures" to distinguish these measures from self-report measures. The study at hand contributes to the development of an experiment-based behavioural task (EBT) whose measures are meant to assess impulsivity.

Within personality questionnaires, where individuals rate their behaviour themselves, impulsive persons are described with adjectives such as spontaneous, hasty, careless, unrestrained, reckless, uncontrollable, irrepressible, acting on the spur of the moment, changeable, impatient, thoughtless, restless, daring, excitable, tempestuous (scale: 'impulsivity', PRF-German version; Stumpf, Angleitner, Wieck, Jackson, & Beloch-Till, 1985). According to Whiteside and Lynam (2001) who integrate several common questionnaire measures of impulsivity into the framework of the Five-Factor-Model FM, impulsive behaviour can be described by four facets of the NEO-PI-R (Costa & McCrae, 1992), namely impulsiveness, excitement seeking, self-discipline and deliberation. The corresponding adjectives to these facet scales include: moody, irritable, sarcastic, pleasure-seeking, daring, adventurous, organized, not lazy, efficient, not hasty, not impulsive, and not careless.

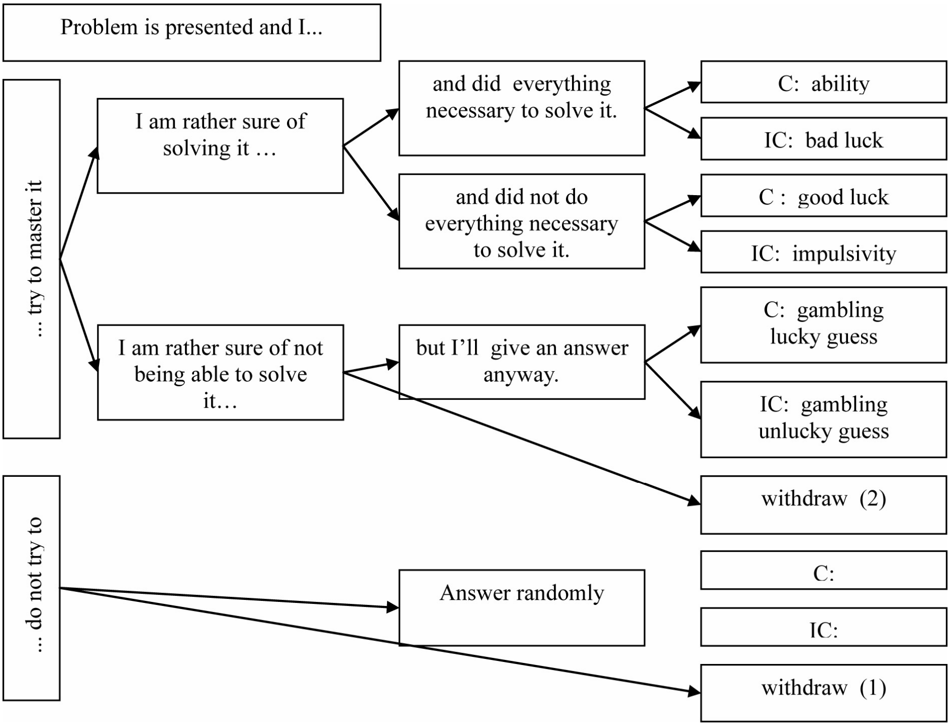
A definition of impulsivity, focusing on the behaviour of impulsive persons, can be found within the cognitive style framework (Gardner, Holzman, Klein, Linton, & Spence, 1959; Messick, 1984). Originally persons with an impulsive style were characterized as persons making quick decisions when faced with a difficult problem (Kagan, 1965), later (Kagan & Kogan, 1970) the quality of the decision was included. Only persons who tend to make quick and incorrect decisions are classified as impulsive.

All of the definitions of impulsivity include the following two elements: (1) Impulsive behaviour includes taking a risk. Both questionnaire based descriptions include the adjective "*daring*". The cognitive style oriented definition refers to a 'difficult problem', here the risk lies in solving the problem or not. Thus the problems presented in style oriented assessment tools are usually rather difficult or time consuming. (2) Impulsive behaviour includes not being careful enough on account of working too quickly. Both questionnaire based descriptions include the adjectives "*hasty*" and "*careless*". The cognitive style oriented definition mentions the "*quick (and wrong) decision*".

In labelling somebody 'impulsive', the cognitive style framework focuses on the output of the behaviour. Only the behaviour leading to an undesirable output (e.g. an incorrect answer) is regarded as impulsive behaviour. Hasty and careless behaviour, not leading to an undesirable output, is not labelled as impulsive behaviour within this theoretic framework. Thus, using the undesirable output as an indicator for impulsivity is a shortcoming of the cognitive styles' definition of impulsivity.

The assessment principle of using a 'difficult problem' for the assessment of impulsivity overlaps with the assessment principle of cognitive ability tests. They too use 'difficult problems' for the assessment. However, the measures used to quantify the respective trait differ – first there are those that are interested in the number of incorrectly answered items, then there are the latter which are interested in the number of correctly answered items. As discussed earlier, in ability testing one is aware of the fact that the score derived from counting correct answered items, not necessarily only reflects ability, but may also contain some lucky guesses. And there have been attempts in ability testing to reduce these unwanted effects on the score. There has also been an interest in discussing whether the incorrect answers on tests assessing impulsivity incorporate impulsive answers only. For the error-score of the traditionally used Matching-Familiar-Figures-Test (MFFT; Kagan, Rosman, Day, Albert, & Philipps, 1964) correlation coefficients between -.51 and -.69 with WISC intelligence scores were reported (Block, Gjerde, & Block, 1986). Beside this, interactions between MFFT score and ability and item difficulty have also been reported, indicating that impulsive persons do not adapt to the speed of working with increasing difficulty (Lawry, Welsh, & Jeffrey, 1983).

A testee confronted with a 'difficult problem' (either within an ability-test, or within an impulsivity test) may make the considerations depicted in figure 1, thus leading us to distinguish four different types of correct answers, as well as four different types of incorrect answers, and two different types of withdrawal. Even when we assume the optimal condition, in other words, that we only have to deal with subjects interested in mastering the problem, three types of correct as well as incorrect and one type of withdrawal have to be dealt with. Thus it can be concluded that correct answers are not necessarily only indicators of



Notes: C = correct, IC = incorrect.

Figure 1:

Considerations leading to correct answers, incorrect answers and withdrawal

ability, and that incorrect answers do not necessarily only show a lack of ability. However, unravelling the meaning of correct as well as incorrect answers may be possible within the framework of the matching law (DeVilliers & Herrnstein, 1976; Herrnstein, 1970; Rachlin, 1971) and the signal-detection theory based experimental research focusing on decision criterion learning (Maddox, Bohil, & Dodd, 2003). Moreover, prospect theory (Kahneman & Tversky, 1988) might also provide us with an interesting insight. The matching law extends the principle of learning theory (Skinner, 1938, 1953) to choice behaviour by interpreting choice as a behaviour set into the context of other behaviour. The frequency of behaviour is a direct function of the frequency of the reinforcement for each behaviour alternative. Assuming that choices are made to maximize gain, behaviour is predictive on basis of the reinforcement. However, it must be noted that although humans' behaviour follows the matching law quite well, there are also effects known that destroy the relation between reinforcement and frequency of behaviour. This finding is not unique to the matching law. Furthermore, signal-detection theory based experimental research results indicate that individuals rarely achieve the optimal decision criterion as estimated by signal detection theory. And prospect theory (Kahneman & Tversky, 1988) states that there is a bigger impact of losses than of gains.

Building on these three approaches we propose the general hypothesis that modifying the situational cues under which the 'difficult task' is presented, helps to unravel the meaning of correct and incorrect responses in a multiple choice task. This will lead to more valid impulsivity scores. It is likely that this also applies to ability scores, but as we will not focus on them we will not elaborate on this part of the hypothesis. It is hypothesized that linking incorrect responses with a loss of points and offering "no decision" as a choice reduces gambling significantly. A score derived from counting incorrect answers will therefore incorporate incorrect answers due to impulsivity and incorrect answers due to bad luck.

The currently most common instrument for assessing impulsivity beyond the questionnaire technique is an EBT using visual discrimination items as stimulus material. The comparing area sizes task (CAS-task; Kubinger & Ebenhöh, 1996; Wagner-Menghin, 2006a) was inspired by subtest T17 from the Objective Test Battery (Häcker, Schmidt, Schwenkmezger, & Utz, 1975), the German version of Cattel's O-A Personality Factor Batteries (Cattell, 1955). It observes testee's behaviour whilst they have to find the larger area size when comparing two amoeba-like shaped figures. The task is administered in a way that encourages the impulsive, spontaneous behaviour of giving a quick answer. It is speeded by showing a count down on the screen above the items which is meant to induce time pressure (30 seconds for 20 pairs). After choosing an answer the next item immediately appears on the screen, at the same time testees are additionally informed that they have to work as quickly as possible. Together with the fact that there are only two shapes to compare, this procedure is meant to induce impulsive persons' feeling of being able to work very quickly and easily through the task. To increase the possibility of errors there are only slight differences between the size of the two areas, making the comparison difficult and time consuming, and one can never be certain that they have solved the problem correctly. There is also an element in the task that encourages reflective behaviour. The option "no decision" when one cannot decide which of the shapes is the larger can be used to switch to the next item. Besides the two variables 'precision' (sum of correct answers/ sum of answers completed; to reduce ability effects) and 'decisiveness' (sum of correct + sum of incorrect answers), an impulsivity vs. reflexivity score is built using a hierarchical combination of the sum of incorrect answers, the sum of correct answers and the sum of "no decision". In accordance with the common practice of focusing on 'undesirable behaviour output' in order to judge whether a subject is impulsive or not, this score is dominated by the sum of incorrect answers – the more wrong answers, the higher the impulsivity score. Persons who answer quickly (and can therefore complete a lot of items within the 30 seconds) and incorrectly are assigned the highest impulsivity score. This score follows the rationale of using undesirable outcome as an indicator for impulsiveness. An undesirable outcome will occur more often if an individual does not spend enough time searching for the right solution. However, there are several aspects which might corrupt the validity of the impulsivity score in the CAS-task. (1) The task is presented as a performance task and individuals are not explicitly informed about the fact that the number of incorrect answers is of relevance for the result. Thus gambling on difficult items might appear a good strategy as when scoring a performance task only correct answers are usually considered, and incorrect answers ignored. Additionally too high a cautiousness will result in a low working speed which also reduces the chance of getting the answers correct. (2) Following this strategy, "no decision" is not attractive as a possibility for answering, which leaves the probability of guessing the right answer as 50 % (as there are only two shapes to compare). The empirical results of Wagner-Menghin and

Reisenhofer (in press), who investigated situational influences on the comparing area sizes task, support this argumentation. They show that within a sample taking part in a personnel selection process, as well as within a sample volunteering in a research project, only about 10 % of the individuals choose "no decision"! Therefore, although the number of incorrect answers is used to measure impulsivity, it is unlikely that this measure actually reflects impulsive behaviour in this context. This situation is unsatisfying as validation studies can not be successful as long as there are no suitable measures.

(3) The CAS-task, which is meant to introduce an emotional challenging condition, is completed under time pressure. However, results obtained within the framework of Gray's neuropsychological model of approach and avoidance learning (Gray, 1987) concerning the breakdown of self-regulation indicate that there are three types of breakdown of self-regulation, each leading to behaviour that is classified as impulsive according to the definitions given at the beginning of this paper. This approach is important as it offers not only a description of impulsive behaviour, but also works on finding an explanation for how this behaviour is generated. Disinhibited behaviour is explained either by a dominance of the behaviour activation system (BAS), which itself is stimulated by the perception of rewards (type I BAS mediated impulsivity), or given the suitable situational cues of not being able to withdraw from a challenging situation, by a dominance of the behaviour inhibition system (BIS), which is stimulated by the perception of punishment (type II anxious impulsivity), or it is described as an intrinsic problem in response modulation (type III response modulation deficit impulsivity) (Newman & Wallace, 1993). Within the CAS-task there is the situational cue activating the BAS – individual's score might increase whenever trying to solve an item – as well as the situational cue of time pressure activating the BIS. Whether this encourages or hinders the development of valid impulsivity measures is currently unclear. (4) Another problem lies in the fact that it is not possible to control how challenging an item is for an individual. The degree of challenge depends on the interaction of a person's ability and an item's difficulty. Given the definition impulsive behaviour occurs '… when facing a challenging situation…' controlling the degree of challenge is thus regarded as central for creating a valid situation for behaviour observation. And as reported above (Lawry, Welsh, & Jeffrey, 1983) differences between impulsive and reflective subjects regarding working speed are only pronounced when working on difficult problems. Therefore it is important to use difficult items.

Within this study we are not able to address all of the described problems. As we currently do not have enough information about the items' difficulties, we are not able to implement an adaptive testing strategy, which can be used to adapt the amount of challenge individually, thus at least controlling the objective challenge between the subjects.

However, we redesigned the described task by enhancing three features to allow for a more valid assessment of impulsive behaviour.

(1) To reduce the chance of guessing a correct answer the answer category *same size* was introduced into the task. This complies with some unsystematic but frequent observations that there are testees that conclude that the areas are of the same size.

(2) We modified the task so as to enable individuals to develop a suitable response strategy. This means that we want to make it obvious that there is a risk in the task, but we also want to make it obvious that it is possible to avoid the risk, which brings us back to the two strategies for dealing with unwanted response styles above. Although unsuccessful in suppressing gambling on ability tests on their own, it is hypothesized here that combining the

"no decision" option approach with the pay-off matrix approach can help to unravel the meaning of correct and incorrect responses in a multiple choice task. Therefore, a payoff matrix defining a gain of points for correct answers and loss of points for incorrect answers is introduced in the instruction of the task. The payoff matrix emphasizes the risk of a decision, and distinguishes the task clearly from the usual – the more correct, the better – ability tasks. According to prospect theory (Kahneman & Tversky, 1988) there is a bigger impact of losses than of gains. Thus it is hypothesized that linking incorrect responses with a loss of points and offering "no decision" as a choice reduces guessing significantly. The matrix also allows defining the meaning of the answer "no decision". In our case the individual has to decide whether to attempt to solve a task or not to solve it in order to maximize the gain from the task. As experiments involving payoff matrices have also been used extensively in signal-detection theory based experimental research focusing on decision criterion learning (Maddox, Bohil, & Dodd, 2003), it is likely that results from this framework might help us to determine the ideal configuration of the payoff matrix. Their results indicate not only that individuals rarely achieve the optimal decision criterion as estimated by signal detection theory, but also that decision criterion learning was better in no-cost conditions (no loss of points in the payoff matrix) in relation to cost conditions (loss of points in the payoff matrix). They also conclude that decision criterion learning was better in cost conditions where there was a long-run loss (LRL) associated with optimal responses than in cost conditions where there was a long-run gain (LRG) associated with optimal responses. This is due to the fact that accuracy maximization was smaller in LRL than in LRG. In order to determine which kind of payoff matrix is ideal for our purpose, results concerning the breakdown of self-regulation within the framework of Gray's neuropsychological model of approach and avoidance learning might again be helpful.

We hypothesize that introducing a payoff matrix that does *not* facilitate the learning of the optimal response (LRG matrices) in the CAS-task will lead to results that differentiate between impulsive individuals and non impulsive individuals. This assumption rests on the results from Newman and Wallace (1993) which show that an impulsive person's behaviour modulation focuses only on part of the situational cues (the potential gain). According to Newman and Wallace, non impulsive persons' behaviour is better modulated as they integrate more situational cues (the potential gain and the loss). To determine which magnitude of costs and benefits is suitable for our purpose, the experimental design will also focus on varying the magnitude of the cost-benefit.

(3) Following the principle of EBT the first trial of the CAS-task is administered without any time limit, which is followed by trials with time pressure.

We have outlined far more problems and possibilities connected with designing an EBT and deriving behaviour measures for impulsivity than we are able to focus on within this study. The main topics of the current research - (a) to study whether the introduced payoff matrix has an effect on the behaviour of the testees and (b) to determine the suitable magnitude of costs and benefits - were chosen because we regard the introduction of the payoff matrix as the key-concept of the task's new design.

Hypotheses: Due to the introduction of the payoff matrix an increase in the usage of "no decision" is expected, independent of the experimental condition. Since Kubinger and Ebenhöh (1996) propose choosing "no decision" as a sign of reflectivity this effect is of interest for a further validation of impulsivity measures. According to learning theory and prospect theory, the higher the loss is for an incorrect answer, the more attractive "no decision" should become, especially when the choice of “no decision” doesn’t cause a loss. All in all, we expect a higher proportion of the sample to use no decision as a choice in comparison to the Wagner-Menghin and Reisenhofer (in press) study (hypothesis a).

An effect on the number of correctly solved items is expected, however two effects are possible. The number of correctly solved items might be lower in groups with a high loss for incorrect responses as testees only solve items they are sure of solving and do not take the chance of occasionally guessing correctly. But there are also arguments that the number of correctly solved items is higher in groups with a high loss for incorrect responses as it is likely that testees are not only more cautious, but will also spend more effort on solving the items (hypothesis b).

One must be aware that introducing the payoff matrix for an ability testing like task is a rather new approach. The instruction is complex and it can be questioned whether testees are able to modify their behaviour according to the payoff matrix introduced (hypothesis c).

Method

To investigate the hypotheses a and b a 2x2 experimental design with the independent variables *loss for incorrect* (low loss = loss for one incorrect answer equals the win for one correct vs. high loss = loss for one incorrect equals the win for two correct) and *loss for no decision* (no loss vs. loss lower than loss for incorrect) was created. According to the findings of the learning theory the ratio of win and loss between the available choices is important. The higher the loss the more infrequent the behaviour is, and the behaviour is also infrequent when other (neutral or rewarded) choices are available. To make guessing less attractive giving an incorrect answer should be risky and result in a loss of points. According to Prospect Theory the value function is shaped asymmetrically, meaning that a loss has a bigger impact than a gain. Therefore, even if one gains or loses the same amount of points the loss is perceived as much worse than the gain is perceived as good. Hence, - given the choice - people would rather prevent the loss of a certain amount of points than try to gain the same amount of points. Therefore, it is hypothesized that in order to make guessing less attractive an incorrect answer should result in a loss of points, and the risk of choosing “no decision” should be less than the risk of giving an incorrect answer so that “no decision” serves as a neutral choice, or at least as a choice with a lower loss. The values for the payoff matrices can be found in table 1. A level of type-I-error $\alpha = 0.05$ is applied.

To investigate hypothesis c the design of hypothesis a and b is expanded to a repeated measurement design where a cross over of the variable “loss for no decision” was added (see again table 1). Again a level of type-I-error $\alpha = 0.05$ is applied.

For a better comparison of the results obtained using our payoff matrices with Maddox et al.’s (2003) results, our payoff matrices are also transformed into the format of the payoff matrix used in their research (table 2). However, although our payoff matrices can be expressed using the terminology of signal-detection theory based research there are two major differences in our research. (1) We are currently not able to keep the base-rate of A and B (probability of the appearance of each true event) constant. However, it should be pointed out that once there is enough information about the difficulty of the items available there is a way to approximate a constant base-rate for each person. In our case the probability of A and B can be interpreted as an interaction of the ability of the person and the difficulty of an item, Thus, using an adaptive-testing algorithm based on the Rasch model should allow us to approximate a base rate of A and B of 50:50. (When the probability of solving an item is the same as not solving an item, choosing A = 'I can decide which area is larger' is as optimal as choosing B = 'I can't decide which area is larger'). (2) We do not have sufficient information concerning the category B true event answers to calculate the SDT measures d' and B.

Table 1:
Experimental design – different payoff matrices

Factor 1	loss for incorrect (independent variable 1)			
	Low		High	
Factor 2	loss for "no decision" (independent variable 2)			
	no loss	lower than for incorrect	no loss	lower than for incorrect
TP1	+2/-2/ 0	+2/-2/ -1	+2/-4/ -0	+2/-4/-3
TP2	+2/-2/ -1	+2 / -2/ 0	+2/ -4/ -3	+2/-4/ -0

Notes: The values in the cells represent the payoff matrices used: correct /incorrect / no decision, TP1 – time point 1, TP2 – time point 2.

Table 2:
Cost-Benefits for each of the experimental conditions using the payoff matrix format of signal-detection theory.

		Category A true event = I can decide which area is larger		Category B true event = I can't decide which area is larger	
		VaA	VbA	VbB	VaB
+2 / -2/ 0	Cost-LRG	+2	-2	0	0
+2/-2/ -1	Cost-LRL	+2	-2	-1	-1
+2/-4/ 0	Cost-LRG	+2	-4	0	0
+2/ -4/ -3	Cost-LRL	+2	-4	-3	-3
+1/ 0/ 0	No cost	+1	0	0	0

Notes: LRG = long run gain, LRL = long run loss (associated with optimal responding); VaA = the value associated with an A response to a Category A true event; VbA = the value associated with a B response to a Category A true event; VbB = the value associated with a B response to a Category B true event; VaB = the value associated with an A response to a Category B true event; +1/0/0 was used with the old task, Maddox, Bohil and Dodd (2003) also classify their matrices indicating the absolute amount of cost and benefit used in the respective matrix as shallow (small amounts) and steep (large amounts). Compared to the amounts used by Maddox, Bohil and Dodd (2003) our amounts are small, thus our matrices are regarded as shallow matrices. Since we have to distinguish between our matrices as well we use the terms "shallow" for the matrices with the smaller values and the term "steep" for the matrices with the larger values.

Instrument

The computerized task (modified CAS-task) was administered in three steps. After introducing the testees to the basic principle behind the task and the gain for correct answers, as well as the loss for incorrect answers and “no decision” option, 30 items were worked on under power conditions without a time limit (power trial; P). This procedure was decided for in order to follow the principles of EBT and to give the testees time to become familiar with the item material, as well as with the payoff matrix. (In the original CAS-task by Kubinger and Ebenhöh (1996) only one speed trial was worked on). Feedback according to the number of points gained from these 30 items could be obtained voluntarily, immediately after finishing the power trial. In step two a different set of items with similar difficulty were administered under time pressure. To increase the reliability, two runs of 30 seconds each were taken (speed trial a = S/a). Step 3 also consisted of two speed trials. To stimulate cognitions about the task and the optimal working strategy, all individuals were informed about their performance after the second and third run; additionally individuals had to predict their performance for the third and the fourth trial (speed trial b = S/b). In order to compare the performance between the three steps, the percentage of correctly solved items was used as a dependent variable. Three variables were composed: *percentage of **c**orrectly solved items in the **p**ower trial* (%correct-P), the aggregated *percentage of **c**orrect answers in speed trials 1 and 2* (%correct-S/a), and also the aggregated percentage of correct answers in speed *trials 3 and 4* (%correct-S/b). Additionally the *percentage of “**no d**ecision” in the **p**ower condition* (%no decision-P), the *percentage of “**no d**ecision” in the aggregated **s**peed trials 1 and 2* (%no decision-S/a) and also the percentage of “no decision” in the aggregated speed trials in *3 and 4* (%no decision-S/b) serve as dependent variables.

Sample

191 subjects (group 1: $n = 46$; group 2: $n = 47$; group 3: $n = 49$, group 4: $n = 49$), aged between 18 and 36 (M = 24.7, SD = 3.99) volunteered in the study. They were recruited by word of mouth at the school for Medical-Technical-Services, and also at the Faculty of Psychology, both located in Vienna (Austria). The study was publicized and captioned ‘The importance of work style for self-regulated learning – are you aware of your working style?’. Participants were offered psychological advice regarding work style in the context of learning and exam or test taking. The advice was based not only on the results of the CAS-task but also a structured interview regarding working style in the context of learning. There was no financial reward for participation.

Results

Hypothesis a: Acceptance of “no decision” as an option for answering. 38 % of the sample never chose “no decision”, 23 % chose it at least once in one of the three trials, 13% in two of the three trials, and 26 % in all three trials. There is no link between the usage of “no decision” and sex ($\chi^2 = 1{,}283$, $p = .733$), as well as no correlation between age and the usage of no decision ($r = -.10$).

Total points obtained: Due to the difficulty of the items a large part of the sample has a negative total of points at the end of most trials, especially in group 3 and 4, where there was a high loss for both incorrect answers and "no decision". As an example, the results of the power trial are reported on: About 50 % of group 1 and 2 and 80 % of group 3 and 4 obtained a negative value. The experimenter was prepared to answer testees' questions regarding this negative value. What is interesting is that although group 1 and 3 would have had the possibility of avoiding a loss of points (no decision was paid off with 0) they obtained a similar amount of points to the comparable groups which had no possibility of avoiding a loss (groups 2 and 4 where no decision paid off with –1 or –3). Descriptive results regarding total points for the power-trial are displayed in table 3.

Hypothesis b: A univariate analysis of variance with the between-subject factors loss for incorrect answers and loss for "no decision" indicates significant differences between the four groups (Levene Test: $F = .742$, $p = .528$; $F = 4{,}522$, $p = .004$) regarding %correct-P. However, only loss for "no decision" accounts for this effect, the results for loss for incorrect answers as well as the interaction are insignificant. Groups experiencing no loss for "no decision" (LRG matrix) obtain less %correct in the power trial than the other groups (LRL matrix).

A related samples-analysis of variance with the within-subject factor *trial* (using the variables %correct-P, %correct-S/a and %correct-S/b) and the between-subject factors loss

Table 3:

Percentage of correct answers, incorrect answers, and "no decision" answers per group

factor 1	loss for incorrect (independent v. 1)							
	low				High			
factor 2	loss for "no decision" (independent v. 2)							
	no loss		lower than for i		no loss		lower than for i	
	group 1 (n = 46) LRG (+2/-2/0)		group 2(n = 47) LRL(+2/-2/-1)		group 3(n = 49) LRG(+2/-4/0)		group 4(n = 49) LRL(+2/-4/-3)	
Variable	M	SD	M	SD	M	SD	M	SD
%correct-P	46	13	**54**	13	48	14	**54**	14
%correct-S/a	49	14	**57**	15	51	15	**53**	14
%correct-S/b	50	12	**52**	14	47	18	**52**	14
%no decision-P	**10**	13	02	05	**09**	11	04	08
%no decision-S/a	**12**	18	02	05	**11**	15	03	08
%no decision-S-b	**11**	17	04	10	**15**	22	03	06
Points Total(P)	**0**	12	5	14	**-22**	17	-21	23

Notes: %correct-P: percentage correct power trial, %correct-S/a, %correct-S/b: percentage correct speed trials aggregated 1 and 2 (a) and 3 and 4 (b). %no decision-P: percentage "no decision" power trial, %no decision-S/a, %no decision-S/b: percentage "no decision" speed trials aggregated 1 and 2 (a) and 3 and 4 (b).

for incorrect answers and loss for no decision is carried out on the data in order to take the full information from all three steps into account. A significant effect of the within-subject factor trial on the dependent variables percent of correct answers emerged (Box-M: $p = .528$, Wilks' Lambda = .963, $F = 3.556$, $p = .031$). The interaction between the within-subject factor trial and both the between subject factors is insignificant (Wilks' Lambda = .991, $F = 0.842$; Wilks'Lambda = 0.982, $F = 1,674$). The interaction trial*loss for "no decision" *loss for incorrect answers just falls short of significance (Wilks' Lambda = 0.971, $F = 2.805$, $p = .063$). Thus it can be said that the %correct-P changes between the three trials and that there is a chance that this change is influenced by the interaction of loss for incorrect answers and loss for "no decision". The descriptive statistics indicate that the shallow LRG-group improves the %correct in the speed trials, whereas the steep LRL groups tend to decrease the %correct. The two other groups show an increase, which is followed by a decrease.

Due to a significant Box-M-Test, for the variables %no decision-P, %no decision-S/a and %no decision-S/b the planned analysis of variances could not be carried out. Kruskal Wallis Tests and Friedman Tests were performed instead (Alpha was adjusted using Bonferoni adjustment). The three Kruskal Wallis Tests for the three dependent variables with the independent variable group yielded significant results (%no decision-P: $\chi^2 = 21.854$; $p<0.001$; %no decision-S/a: $\chi^2 = 22.572$, $p<0.001$, %no decision-S/b: $\chi^2 = 19.949$, $p<0.001$), with the highest mean rank for group 3, followed by the groups 1, 4 and 2. This meets the expectations. For group 1 (shallow LRG) and 3 (steep LRG) the choice of "no decision" does not cause any loss, therefore this alternative is relatively more attractive than for the groups 2 (shallow LRL) and 4 (steep LRL). Although the loss is smaller when both groups use "no decision" than when they answer incorrectly, the relative loss when using "no decision" is higher for group 2 than for group 4 - using "no decision" should therefore be more attractive for group 4 than for group 2. The mean ranks reflect these relations.

The four Friedman tests for the four experimental groups using %no decision-P, %no decision-S/a and %no decision-S/b to form the within factor trial yielded insignificant results (Group 1: $\chi^2 = 0.637$, $p = 0.727$; group 2: $\chi^2 = 3.351$, $p = 0.187$; group 3: $\chi^2 = 0.514$, $p = 0.773$; group 4: $\chi^2 = 5.871$, $p = 0.053$).

In summarizing these results one can say that there is an effect of the experimental manipulation on the ratio of choosing "no decision", but there is no effect between the three steps of the procedure. The ratio of "no decision" stays more or less constant between the power and the speed trials. The expected relations between the four groups could be found.

Hypothesis c: Using repeated measurement ANOVA (for the variables %correct-P-TP1 and %correct-P-TP2) it was found that there is no significant effect of the independent variable time point (Wilks' Lambda = .993, F = 1.411, p = .236) but that there is a significant interaction effect between the variable time point and the variable group (Wilks' Lambda = .823, F = 13.429, p<.001). This means that the overall performance regarding %correct-P between the two time points is the same, but the performance per group changes. Those groups that achieve higher %correct-P at time point 1 achieve less at time point 2, and those that achieve less at time point 1 have more at time point 2. These results indicate that the individuals adapt their behaviour according to the payoff matrix.

Discussion

The experiment contributes to developing an EBT focussing on response style with a multiple-choice ability task to derive behavioural measures for the assessment of impulsivity. In order to learn more about how to influence the response behaviour of an individual in such a task, a payoff matrix together with ‘no decision’ as a choice was introduced in the CAS-task, a task currently in use for the assessment of impulsivity. The combination of these formerly less successful attempts to influence the response style was found to have several effects on the variables %no decision and %correct derived from the CAS-task. Additionally, the former speeded CAS-task was expanded, consisting now of a-non speeded power trial followed by four speed trials, thus following the principles of EBTs.

The aim of making “no decision” a seriously attractive alternative in the CAS-task – by introducing the payoff matrix - was achieved. Only 38 % of the sample never used “no decision” as a choice, compared to the 90 % who never used “no decision” with the original CAS-task mentioned in the introduction (Wagner-Menghin & Reisenhofer, in press); this is considered a relevant increase.

But it should also be noted that the experimental manipulation has an effect on the usage of “no decision”. Within groups with LRG “no decision” is chosen more often than within groups with LRL. So when faced with the chance of possibly loosing -2 (or -4) points or definitely loosing -1 (or -3) points individuals are more inclined to take the risk of loosing the points, with a chance of gaining some, instead of accepting a minor loss without the chance of a possible gain. Of great interest in this context is the observation that within the LRL groups the %correct is higher than in the LRG groups and that these groups earn more total points in the power trial. This is in accordance with the results from Maddox et al. (2003) which conclude, that LRL leads to a better learning of the optimal decision criterion – here the optimal strategy to maximize the gain. However, it contradicts our initial assumption of an influence due to the amount of loss for incorrect responses. Instead our results indicate that the %correct is only influenced when “no decision” is connected with a loss, whereas the amount of loss for incorrect answers has no influence on the %correct. Whether this is a general effect or due to insufficient experimental manipulation (e.g. difference between low and high too small) is currently unsolved. In some way our results can also be explained by the prospect theory, which states that in the case that loss is more likely than gain, individuals tend to make risky decisions, as opposed to conditions where gain is more likely where individuals tend to make safe decisions. This is called the reflection effect (Tversky & Kahneman, 1981). As “no decision” resulted in a loss of points in the LRL condition, testees perceived their chances of losing points as greater and therefore had a tendency towards risky decisions, meaning they would rather take a guess than choose “no decision”.

To fully understand the effect of the payoff matrix on the %no decision, let us return to our initial assumption on how the payoff matrix will influence the %correct. We hypothesized that in groups with a high loss (for incorrect answers) testees will either have more correct answers due to being more cautious and concentrated on the task, or less correct answers because they miss out on a few lucky guesses as a result of not risking a guess. In contrast to our assumption our results indicate that it is not the possible loss of points for incorrect answers, but rather the loss or non loss when choosing “no decision” that influences the %correct. Individuals in the LRL condition achieve more correct answers than others. As their choice of "no decision" is lower compared to the others, the higher %correct is interpreted as the result of being more concentrated on the task instead of being more cautious. Following this argumentation the choice of "no decision" has to be interpreted, not as an act of cautiousness, but as a lack of effort on the task, which in turn leaves us with some doubt that "no decision" indicates a person's reflectiveness as suggested by Kubinger and Ebenhöh (1996). Choosing "no decision" might rather express the tendency to withdraw from the challenging task; in that case it represents type II-anxious impulsivity behaviour. Individuals who have more "no decisions" are therefore expected to answer quickly and incorrectly when withdrawal is impossible. Again the results from the LRL condition are of interest here. Although the LRL can be interpreted as "withdrawal is impossible", the %correct is higher in the LRL condition than in the LRG conditions, which contradicts the assumption of a higher number of quick and incorrect answers due to anxious impulsive reactions.

What do we learn about optimizing our task for the assessment of impulsivity? As expected the payoff matrix with LRG puts more emphasis on incorrect answers and offers a possibility of avoiding the risk of incorrect answers without the loss of points. The results concerning the LRL condition are interpreted as suitable for stimulating the self-control-mechanism, as withdrawing here is not as attractive as in the LRG condition and the %correct is higher than in the LRL condition.

Of additional interest is that the %no decision at the group level stays rather constant between the three trials of the instrument, although it must be mentioned that there are indications of an interaction between the steepness and the type of the payoff matrix. In the shallow LRG matrix the %correct is improved over the three trials. However, unless some of the problems that are still connected with the task are solved (e.g. controlling the item difficulty individually), one should not put too much weight on this result.

Of encouragement, under the aspect of further validation of an impulsivity measure from the CAS-task, is that the individuals used the situational cues from the payoff matrix to adapt their behaviour between the two time points. Under conditions where choosing no decision definitely meant avoiding a loss, it was chosen more often. Thus using a payoff matrix together with the "no decision" option is a suitable approach to influence response behaviour.

As mentioned before one possible limitation of the study is the fact that it has not been possible to control subjects' ability. This should especially be noted, as the items of the task, although tested in a pilot study, turned out to be extremely difficult for the sample. For further studies and for the calibration of an item pool suitable for adaptive testing we recommend adding easier items. It may be helpful to also use various item content or various multiple-choice formats with the LRG payoff matrices in order to further understand the process leading up to choosing "no decision". In summarizing our conclusions we suggest testing whether our results can be replicated using an adaptive procedure before discussing whether to cease using speeded trials as part of the standardized EBT procedure, and introducing a systematic variation of payoff as part of the standardized EBT procedure, as we feel this would be helpful for further developing measures to assess impulsivity.

Acknowledgments

We would like to thank the reviewers for valuable suggestions regarding the decision theory. Thanks also to Nadja Weltzien who worked on discussing the results with respect to the prospect theory during her internship in our department.

References

Amthauer, R. (1972). *Praktisch technischer Verständnistest. [Practical and Technical Comprehension.]*. Göttingen, Germany: Hogrefe.

Block, J., Gjerde, P. F., & Block, H. (1986). More misgivings about the Matching Familiar Figures Test as a measure of reflection - impulsivity: Absence of construct validity in preadolescence. *Developmental Psychology, 22*, 820-831.

Cattell, R. B. (1955). *The O-A (Objective-Analytic) Personality Factor Batteries: Adult and Child Forms.* Champaign, IL: IPAT.

Cattell, R. B. (1958). What is "objective" in "objective" personality tests. *Journal of Counseling Psychology, 5*, 285-289.

Cattell, R. B., & Warburton, F. W. (1967). *Objective personality and motivation tests. A theoretical introduction and practical compendium.* Chicago: University of Chicago Press.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurements (4th ed.)*. Mountain View, CA: Mayfield.

Costa, P. T. J., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

DeVilliers, P. A., & Herrnstein, R. J. (1976). Toward a law of response strength. *Psychological Bulletin 83*, 1131-1153.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H., & Spence, D. S. (1959). Cognitive Control: A study of individual consistencies in cognitive behaviour. *Psychological Issues, 1*(4), 1-186.

Gittler, G. (1990). *Dreidimensionaler Würfeltest. [Three-Dimensional Cube Comparison Test (3DC)]*. Weinheim, Germany: Beltz.

Gray, J. A. (1987). Perspectives on anxiety and impulsivity. *Journal of Research in Personality, 21*, 493-509.

Häcker, H., Schmidt, L. R., Schwenkmezger, P., & Utz, H. E. (1975). *Objektive Testbatterie, OA-TB 75 [Objective Test Battery 75]*. Weinheim, Germany: Beltz.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243-266.

Kagan, J. (1965). Individual differences in the resolution of response uncertainty. *Journal of Personality and Social Psychology, 2*, 154-160.

Kagan, J., & Kogan, N. (1970). Individual variation in cognitive process. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, pp. 1273-1365). New York: Wiley.

Kagan, J., Rosman, B. L., Day, D., Albert, J., & Philipps, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs, 78*, 1-37.

Kahneman, D., & Tversky, A. (1988). Prospect theory: An analysis of decision under risk. In P. Gardenfors & N.-E. Sahlin (Eds.), *Decision, probability, and utility: Selected readings* (pp. 183-214). New York: Cambridge University Press.

Kubinger, K. D. (2006). Ein Update der Definition von Objektiven Persönlichkeitstests: Experimentalpsychologische Verhaltensdiagnostik. [Updating the definition of objective personality tests: experiment based behavioural assessment]. In T. M. Ortner, R. T. Proyer & K. D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests. [Theory and Practice of Objektive Personality Tests]* (pp. 38-52). Bern, Swizerland: Hans Huber.

Kubinger, K. D., & Ebenhöh, J. (1996). *Arbeitshaltungen. Kurze Testbatterie: Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation, Reflexivität/Impulsivität. [Work Styles. Short test battery: Aspiration level, resistance to frustration, achievement motivation, reflexivity/impulsivity. English Version 2004: In: The Vienna test system. Mödling: Dr. G. Schuhfried. GmbH.]*. Frankfurt/M, Germany: Swets.

Lawry, J. A., Welsh, M., & Jeffrey, W. E. (1983). Cognitive tempo and complex problem solving. *Child Development, 54*, 912-920.

Maddox, W. T., Bohil, C. J., & Dodd, J. L. (2003). Linear Transformations of the Payoff Matrix and Decision Criterion Learning in Perceptual Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1174-1193.

Messick, S. (1984). The nature of cognitive styles: Problems and promises in educational practice. *Educational Psychologist, 19*, 59-74.

Newman, J. P., & Wallace, J. F. (1993). Diverse pathways to deficient self-regulation: Implications for disinhibitory psychopathology in children. *Clinical Psychology Review, 13*, 699-720.

Ortner, T. M., Proyer, R. T., & Kubinger, K. D. (Eds.). (2006). *Theorie und Praxis Objektiver Persönlichkeitstests. [Theory and Practice of Objective Personality Tests]*. Bern, Swizerland: Hans Huber.

Rachlin, H. C. (1971). On the tautology of the matching law. *Journal of the Experimental Analysis of Behavior, 15*, 249-251.

Skinner, B. F. (1938). *The behavior of the organism*. New York: Appleton.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Stumpf, H., Angleitner, A., Wieck, T., Jackson, D. N., & Beloch-Till, H. (1985). *Personality Research Form Deutsche Version. [Personality Research Form - German version]*. Göttingen, Germany: Hogrefe.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453 - 458.

Wagner-Menghin, M. M. (2006a). "Flächengrößen vergleichen": Ein Objektiver Persönlichkeitstest zur Erfassung des Stilmerkmals "Impulsivität". ['Comparing shapes': An Objective Personality Test for the assessment of the style impulsivity.] In T. M. Ortner, R. Proyer & K. D. Kubinger (Eds.), *Theorie und Praxis Objektiver Persönlichkeitstests. [Theorie and Practice of Objective Personality Tests]*. (pp. 192-201). Bern (Swizerland): Huber.

Wagner-Menghin, M. M. (2006b, July 6-8). *Objective Personality Tests for Psychological Assessment.* Paper presented at the ITC 5th International Conference on Psychological and Educational Test Adaptation across Language and Cultures, Brussels.

Wagner-Menghin, M. M., & Reisenhofer, S. (in press). Ein OPT zur Erfassung von Impulsivität: Situationsabhängigkeit von verhaltensbasierten Kennwerten; Beziehung zu den BIG 5. [An objective personality test assessing impulsivity: situational influences of behaviour-based measures, relation to facets of the Big-5]. In W. Sarges & D. Scheffer (Eds.), *voraussichtlich: Innovative Ansätze in der Eignungsdiagnostik. [expected: Innovations in career assessment]*.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*, 669-689.