Chapter 1

Typical scoring of implicit measures

This chapter presents typical scoring procedures for IAT data and SC-IAT data, as well as the development of new, open source tools for easily computing IAT and SC-IAT scores.

The typical scoring algorithms for both the IAT and the SC-IAT are illustrated and described. It is not unusual to find studies in which both the IAT and the SC-IAT have been administered together, and their performances, for example for predicting a behavioral outcome, have been compared. However, this comparison might be biased by many differences concerning both the administration and scoring procedures of the two implicit measures. Therefore, new scoring algorithms have been introduced with the aim of reducing the noise due to external factors (i.e., the scoring procedure itself) in the comparison between implicit measures. The new scoring algorithms are presented. The results of an empirical study in which the performance of typical and modified scoring algorithms have been compared in respect to the prediction of a behavioral outcome are reported.

The core computation of both scores is rather easy, but the many steps that have to be

undertaken for preparing and cleaning the data makes it an error-prone procedure, hence hindering the reproducibility of the results. Since there is a lack of easy-to-use and open source tools for their computation, a Shiny app and an R package have been developed for the computation the IAT and the SC-IAT D scores. These tools are presented at the end of this chapter.

1.1   IAT D score

? (?) introduced different variations of the D score algorithm (Table 1.1). These variations result from the combination of error correction strategies ("Error replacement" in the Table) and treatment for fast responses ("Lower tail treatment" in the Table).

Table 1.1: Overview of D score algorithms.

Algorithm	Error replacement	Lower tail treatment
D1	Built-in correction	No
D2	Built-in correction	Delete trials < 400 ms
D3	Mean $+ 2sd$	No
D4	Mean $+ 600$ ms	No
D5	Mean $+ 2sd$	Delete trials < 400 ms
D6	Mean $+ 600$ ms	Delete trials < 400 ms

Note: For all algorithms, trials with latency $> 10,000$ ms are discarded. For the algorithm in which error responses are replaced with the average response time plus a penalty, the average response time is computed on correct responses only.

Blocks B1, B2, and B5 in Table **??** are considered as pure practice blocks and are hence

discarded from the computation. Only trials from Blocks B3, B4 (i.e., Mapping A) and B6, B7 (i.e., Mapping B) are used for the computation.

The error correction strategies based on built-in correction (*D1* and *D2*) refer to the IAT procedure including feedback, according to which respondents have to correct their error responses. The response time considered for the computation of the *D* score is the response time at the first (incorrect) response inflated by the time required to correct it. All other algorithms (from *D1* to *D6* in Table 1.1) use a *post-hoc* error correction strategy, for which error responses are replaced by the average response time of the correct responses in the block in which the error occurred increased by a standard penalty (i.e., either 600 ms or twice the standard deviation). The other feature differentiating the *D* score algorithms is the lower tail treatment, according to which fast trials (trials faster than 400 ms) are discarded or not.

Regardless of the specific features of each algorithm, the core procedure for computing the *D* score is the same. Firstly, the *D* scores of associative practice blocks (Eq. 1.1):

$$D_{\text{practice}} = \frac{M_{\text{B6}} - M_{\text{B3}}}{SD_{\text{B6, B3}}}, \tag{1.1}$$

and of associative test blocks (Eq. 1.2)

$$D_{\text{test}} = \frac{M_{\text{B6}} - M_{\text{B3}}}{SD_{\text{B6, B3}}}, \tag{1.2}$$

ar computed. In both cases, the difference in the average response time between the two critical blocks is divided by the standard deviation computed on the pooled trials of both

blocks. Once D_{practice} and D_{test} are obtained, it is possible to compute the actual D score:

$$D \text{ score} = \frac{D_{\text{practice}} + D_{\text{test}}}{2}. \tag{1.3}$$

The block order in Equation 1.1 and Equation 1.2 is arbitrary and can be reversed. The resulting D has to be interpreted accordingly. In the Coke-Pepsi IAT in Chapter **??**, Blocks B3 and B4 constituted the Coke-Good/Pepsi-Bad condition. Conversely, the Pepsi-Good/Coke-Bad condition was Blocks B6 and B7. If the D score is computed following the order of the blocks in Equation 1.1 (i.e., $M_{\text{B6}} - M_{\text{B3}}$) and in Equation 1.2 (i.e., $M_{\text{B7}} - M_{\text{B4}}$), a positive score would indicate slower responses in Pepsi/Good-Coke/Bad condition than in Coke/Good-Pepsi/Bad one, probably indicating a preference for Coke over Pepsi. Vice versa, if the order of the Block in Equations 1.1 and 1.2 is reversed (i.e., $M_{\text{B3}} - M_{\text{B6}}$ and $M_{\text{B4}} - M_{\text{B7}}$, respectively), a positive score would indicate slower responses in the Coke/Good-Pepsi/Bad condition than in the Pepsi/Good-Coke/Bad condition, indicating a possible preference for Pepsi over Coke.

1.2   SC-IAT D score

Since blocks B1 and B3 (Table **??**) are considered as pure practice blocks, they are discarded from the computation of the SC-IAT D score. If a rtw was included in the administration procedure, all responses exceeding it are considered as non-responses and are discarded from the computation. All responses with a latency faster than 350 ms are discarded, and error

responses are replaced with the average response time of the block in which the error occurred inflated by a standard penalty of 400 ms.

After cleaning and preparing the data, the SC-IAT *D* score is simply computed as the difference in the average response time of the two critical blocks (i.e., $M_{B4} - M_{B2}$) divided by the standard deviation computed of the correct trials on both blocks. As for the IAT, the order of the critical blocks is arbitrary and the interpretation of the *D* score changes accordingly.

In the Coke SC-IAT example illustrated in Chapter **??**, Block B2 was the Coke-Good condition, while Block B4 was the Coke-Bad condition. Following this structure, if the *D* score is computed as the difference between Blocks B4 and B2, a positive score would indicate slower responses in the Bad/Coke condition than in the Good/Coke one, standing for a positive evaluation of Coke. Vice versa, if the score is computed in the opposite direction (i.e., $M_{B2} - M_{B4}$), a positive score would indicate slower responses in the Good/Coke condition than in the Bad/Coke condition, indicating a plausible negative evaluation of Coke.

1.3   A fairer comparison between the IAT and the SC-IAT

In Study 1, ? (?) directly investigated and compared the predictive ability of a Coke-Pepsi IAT, that of a Coke SC-IAT, and that of a Pepsi SC-IAT, in respect to a soda. As their results suggested, measures obtained from both the Coke-Pepsi IAT and the Pepsi SC-IAT played a role in predicting the actual soda choice, while the measure obtained from the Coke SC-IAT did not contribute to the choice prediction. Drawing on these results, authors speculated that the soda choice is more guided by a positive evaluation of Pepsi than by a negative

evaluation of Coke. Nonetheless, the direct comparison between the predictive ability of the two implicit measures has been poorly investigated. Despite the study by ? (?) provided interesting information on the functioning of implicit processes and the comparison between implicit measures, it also had some shortcomings that might have undermined the validity of their results. The aim of the study reported in this section was to to provide a fairer comparison of the predictive ability of the IAT and the SC-IAT in respect to a behavioral choice by presenting new scoring algorithms for the two implicit measures. This study is published in ? (?).

Among the shortcomings in ? (?), the sample size was rather small, and results should hence be interpreted with caution. Moreover, the comparison between the predictive ability of the implicit measures might have been affected by issues concerning their administration and scoring procedures. The IAT and the SC-IAT differed in the number of trials, the number of blocks, and the number of exemplars representing each category. The SC-IAT employed more trials and more stimuli than the IAT, for both the evaluative dimensions (twenty-one exemplars for each SC-IAT evaluative dimension versus five exemplars for each IAT evaluative dimension) and the object categories (seven exemplars for each SC-IAT target object category and five exemplars for each IAT target object category). Furthermore, the administration of the SC-IAT included a rtw, while that of the IAT did not have such a constraint on the responses. The presence of a rtw makes the task more difficult and produces a sense of urgency that is otherwise missing (?, ?). Additionally, in the SC-IAT respondents were given feedback for each correct and incorrect response, while the IAT administration procedure did not include any feedback. The labels used for representing the positive and negative evalua-

tive dimensions changed across implicit measures (*Pleasant* vs *Unpleasant* for the IAT and *Good* vs *Bad* for the SC-IAT), as well as the response keys used for sorting the stimuli. The IAT *D* scores were computed according to the *D* score procedure in ? (?), despite ? (?) failed to report the exact algorithm they employed. The SC-IAT *D* score presented in Section 1.2 was used for computing the SC-IAT *D* score.

Given the differences between administration and scoring, the comparison between the ability of the IAT and that of the SC-IAT to predict a behavioral outcome might have been unfair. To the best of our knowledge, there is neither a scoring procedure employing the same criteria for both the IAT and the SC-IAT, nor an attempt to align the two implicit procedures to allow for a fairer comparison between their predictive capacity. It would be interesting to compare the predictive ability of the two implicit measures by using the same scoring procedure on their data and by keeping the administration as similar as possible, while acknowledging their key features (e.g., block types and usual length of the blocks). If by using the same scoring procedure and by reducing the administration-related differences there are still differences in the predictive ability of the two measures, these differences can be reasonably attributed to the implicit procedure itself.

To obtain a fairer comparison between the two implicit measures, both administration (e.g., stimuli, response time window, feedback) and scoring of the two procedures have been aligned.

1.3.1   Method

To test the predictive ability of the new scoring procedures, one Chocolate IAT, one Dark chocolate SC-IAT, and one Milk chocolate SC-IAT were developed. The decision to use chocolate as the object category was driven by different reasons. Firstly, chocolate preference should not be sensitive to social desirability, and hence respondents would have no concerns in explicitly reporting their actual chocolate preference. Moreover, it offers the chance to ask for a behavioral choice disguised as a reward for the participation.

Inquisit 3.0 (?, ?) was used for administering the implicit measures (i.e., the IAT and the two SC-IATs) and the demographic questionnaire.

Participants.   Participants were recruited at the University of Padova. One-hundred and sixty-one people (F $= 63.55\%$, Age $= 23.95 \pm 2.83$) volunteered to take part in the study, with no compensation. Participants were informed about the confidentiality of the data, and the possibility of withdrawing from the experiment at any time they wished. They were asked for their consent to take part in the study. Majority of the participants were students (94.08%), including both undergraduates, master, and PhD students. Only two participants reported to have a PhD title, while the majority reported to have a bachelor's degree (43.42%), immediately followed by those who reported having a high school diploma (32.24%) and a master's degree (23.03%).

Materials and Procedure.   Chocolate stimuli were composed by seven images of chocolate that were modified to represent either Dark or Milk chocolate. Seven images for each type

of chocolate were used. Three independent judges evaluated the stimuli regarding their properties, specifically whether they were clearly identifiable as dark or milk chocolate images. The three judges agreed on the representativeness of the stimulus of the category to which it was supposed to belong. All chocolate images were presented on a white background.

In the Chocolate IAT, both Dark and Milk chocolate images were used. In the two SC-IATs, only either dark (Dark SC-IAT) or milk (Milk SC-IAT) chocolate images were used. Object categories were labeled *Dark* or *Milk*. Evaluative attributes categories were composed of 13 stimuli each. The evaluative categories were labeled as *Positive* (i.e., "good", "laughter", "pleasure", "glory", "peace", "happiness", "joy", "love", "wonderful", "beautiful", "excellent", "heaven", "marvelous") or *Negative* (i.e., "evil", "bad", "horrible", "terrible", "annoying", "pain", "failure", "hate", "nasty", "disaster", "agony", "ugly", "disgust"). Response key "E" was used for sorting the stimuli belonging to the categories represented on the left-side of the screen. Response key "I" was used for sorting the stimuli belonging to the categories represented on the right side of the screen. The SC-IAT practice Blocks B1 and B3 were composed of 20 trials, as for the practice blocks of the IAT. Neither the IAT nor the SC-IATs included any feedback or rtw. Respondents were asked to be as fast and as accurate as they could in performing the tasks.

Respondents were explicitly asked to report their evaluation for Dark and Milk chocolate on two distinct items ("How much do you like Dark/Milk chocolate?") rated from "0 – Not at all" to "5 – Very much". The order of presentation of the implicit measures was counterbalanced across participants, and the explicit assessment was kept constant after the administration of implicit measures.

As a reward for their participation, respondents were offered with either a free Dark chocolate bar or a Milk chocolate one. The experimenter registered respondents' choices after they left the laboratory.

Chocolate IAT: The critical blocks were composed of 60 trials each (20 practice and 40 test), defining the Dark-Good/Milk-Bad condition (DGMB), and the Milk-Good/Dark-Bad condition (MGDB).

Dark SC-IAT: The critical blocks were composed of 72 trials each, defining the Dark-Good/Bad (DG) and the Good/Dark-Bad (DB) conditions.

Milk SC-IAT: As for the Dark chocolate SC-IAT, the critical blocks were composed of 72 trials each, defining the Milk-Good/Bad (MG) and the Good/Milk-Bad (MB) conditions.

1.3.2   Data analysis

Data cleaning and *D* score

All IAT D scores not including a built-in correction (algorithms $D3$, $D4$, $D5$ and $D6$ in Table 1.1) were computed. The procedure described in Section 1.2 was followed for computing the SC-IAT D score.

The scoring algorithms that have been introduced in this study result from different combinations of two main characteristics, namely the trials on which the standard deviation for the replacement of error responses was computed (i.e., only correct trials or all trials) and the quantity used for standardizing the difference in the response times between conditions (i.e., Cohen's pooled standard deviation or pooled trials standard deviation). Cohen's pooled

standard deviation for two groups of size n_1 and n_2 is as follow:

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1 + (n_2 - 1)SD_2}{n_1 + n_2 - 2}} \tag{1.4}$$

The resulting eight combinations (identified by letter "*m*", *modified*) are illustrated in Table 1.2.

Table 1.2: Overview of modified algorithms. for computing IAT and SC-IAT scores.

Features	*m*1	*m*2	*m*3	*m*4	*m*5	*m*6	*m*7	*m*8
Lower tail treatment				< 350 ms				
Upper tail treatment				> 10,000 ms				
Error treatment	Mean (Correct) + 2 *sd* (Correct)				Mean (Correct) + 2 *sd*			
Denominator	Pooled trials		Cohen		Pooled trials		Cohen	
Denominator trials	Correct	All	Correct	All	Correct	All	Correct	All

While the typical procedure for the SC-IAT includes a default lower tail treatment, the lower tail treatment for the IAT depends on the *D* score algorithm (see Table 1.1). To have a comparable score, a common lower tail treatment for both procedures was set (i.e., responses with a latency less than 350 ms were discarded). Since it is not uncommon to find SC-IATs with no rtw, a common upper tail treatment for response times was proposed for both implicit measures (i.e., responses over 10,000 ms were discarded). Concerning the SC-IAT upper tail treatment, it might be argued that the deletion of the responses higher than 1,500 ms (i.e., the rtw cut-off) could be a more appropriate threshold for slow responses. Nonetheless, the presence of the rtw itself produce an urge to respond that is missing when the rtw is not

included in the administration procedure (?, ?).

Since the SC-IAT is known to be an easier task than the IAT (?, ?), the latency of the responses in the SC-IAT tend to be faster than the latency of the responses in the IAT. Therefore, assuming 600 ms as a reasonable time for correcting the error response might be a too strong assumption for the SC-IAT data. Conversely, the penalty used in the SC-IAT (400 ms) might be not enough for acknowledging the response time needed for correcting the error response in the IAT. For this reason, the error responses are replaced by the average response times in the block in which the error occurred inflated by two times the standard deviation of the block.

The pooled trials standard deviation and Cohen's pooled standard deviation were computed either considering only correct responses or all trials. In the former case, the variability due to incorrect responses is not accounted for, while it is addressed in the latter case. Finally, the IAT modified procedures were computed as the difference between the two associative conditions, instead of as the mean of the standardized average response time differences between the practice and test blocks.

Both typical and modified scoring for the IAT were computed so that positive scores indicated faster responses in associating Milk chocolate with positive attributes and Dark chocolate with negative attributes, and hence an implicit preference for Milk chocolate over Dark chocolate. Conversely, negative scores indicated faster responses in associating Dark chocolate with positive attributes and Milk chocolate with negative attributes, indicating an implicit preference for Dark chocolate over Milk chocolate.

For the SC-IATs, both typical and modified procedures were computed so that positive

scores indicated faster responses in associating the target chocolate with positive attributes than with negative attributes, hence indicating an implicit positive evaluation of the target chocolate. Conversely, negative scores indicated faster responses when the target chocolate was associated with negative attributes, and an implicit negative evaluation of the target chocolate.

Consistency between modified and typical scores, and relationship with explicit measures

Pearson's correlations between explicit chocolate evaluations, typical and modified scoring were computed.

Moreover, Pearson's correlations were computed between the typical and modified scores to check for their consistency.

Prediction of the behavioral outcome

Both typical and modified scores were regressed on the chocolate choice, coded as 0 for Dark chocolate choice (DCC) and 1 for Milk chocolate choice (MCC). Each score was regressed on the choice in a separate logistic regression. Since the choice was presented as a dichotomous task in which Dark chocolate was contrasted with Milk chocolate, it is plausible that the relative preference for one chocolate over the other played a role in determining the actual choice. Each SC-IATs score conveys a unique information regarding the absolute positive or negative evaluation of one type of chocolate, therefore lacking a part of information that might be crucial in determining the choice. Using the linear combination of both

SC-IATs scores or a combined score might solve this issue. However, since SC-IATs scores are obtained from two different experiments, their combination, either linear or in a comprehensive score, might be considered a stretch. Grounding on these considerations, both single Dark SC-IAT and Milk SC-IAT scores, as well as their linear combinations, were used for predicting the choice.

Nagelkerke's R^2 (?, ?) and model accuracy of prediction (?, ?) were used as criteria for investigating the scores best accounting for the behavioral choice. Specifically, model general accuracy (i.e., ratio between the number of chocolate choices correctly identified by the model and the total number of choices), DCC accuracy (i.e., ratio between the number of DCCs correctly identified by the model and the total number of observed DCCs), and MCC accuracy (i.e., ratio between the number of MCCs correctly identified by the model and the total number of observed MCCs) were computed.

1.3.3 Results

Data from nine participants were discarded. Eight of them explicitly reported not understanding the tasks they were asked to perform in either the IAT or one of the SC-IATs, while one of them registered too many fast responses, specifically on the Dark chocolate SC-IAT (more than 30% of responses with a latency lower than 350 ms). The final sample was composed of 152 participants (F = 63.82%, Age = 24.03 ± 2.82). Milk chocolate was chosen by the 48.03% of the participants.

The median for the explicit evaluation of Dark chocolate was 3 (Q_1 = 2, Q_3 = 5). The median for the explicit evaluation of Milk chocolate was 4 (Q_1 = 3, Q_3 = 4). No trials

exceeding the threshold of 10,000 ms were found in the SC-IATs. Three trials exceeding the 10,000 ms threshold were found in the IAT, and they were eliminated. The lowest percentage of trials faster than both 400 ms (1.39%) and 350 ms (0.19 %) were found in the IAT. The two SC-IATs showed similar percentages of trials faster that 350 ms (1.00% and 0.90% in Milk SC-IAT and Dark SC-IAT, respectively), as well as of trials faster than 400 ms (4.40% and 4.32% in Dark SC-IAT and in Milk SC-IAT, respectively). All implicit measures had the same overall percentage of correct responses (95%).

In the IAT, the overall average response time was 862.03 ms ($sd = 496.50$, *skewness* $= 3.45$, *kurtosis* $= 22.01$). The average response time in the DGMB condition was 976.44 ms ($sd = 555.19$, *skewness* $= 2.88$, *kurtosis* $= 14.01$) and that in the MGDB condition was 747.62 ms ($sd = 398.30$, *skewness* $= 4.76$, *kurtosis* $= 48.62$).

The overall average response time in the Dark SC-IAT was 679.45 ms ($sd = 328.72$, *skewness* $= 4.10$, *kurtosis* $= 27.94$). The average response time in the DB condition was 673.71 ms ($sd = 322.87$, *skewness* $= 3.90$, *kurtosis* $= 24.46$) and that in the DG condition was 685.19 ms ($sd = 334.39$, *skewness* $= 4.27$, *kurtosis* $= 30.86$).

The overall average response time in the Milk SC-IAT was 675.90 ms ($sd = 322.31$, *skewness* $= 4.48$, *kurtosis* $= 38.19$). The average response time in the MB condition was 695.72 ms ($sd = 344.84$, *skewness* $= 4.10$, *kurtosis* $= 28.32$) and that in the MG condition was 656.08 ms ($sd = 296.78$, *skewness* $= 4.98$, *kurtosis* $= 54.05$).

Relationship with explicit measures. Descriptive statistics for the typical scoring of all implicit measures, along with their correlation with explicit measures, are reported in Ta-

ble 1.3.

Table 1.3: Descriptive statistics of the scores and correlations (r) with explicit chocolate evaluations.

	Modified	M (sd)	Min	Max	r_{Milk}	r_{Dark}	Typical	M (sd)	Min	Max	r_{Milk}	r_{Dark}
IAT	m1	0.64 (0.62)	−1.91	1.72	0.40***	−0.36***	D3	0.41 (0.41)	−1.29	1.25	0.42***	−0.38***
	m2	0.64 (0.60)	−1.86	1.69	0.40***	−0.37***	D4	0.39 (0.39)	−1.26	1.27	0.41***	−0.38***
	m3	0.73 (0.73)	−2.25	2.84	0.39***	−0.34***	D5	0.40 (0.41)	−1.29	1.29	0.42***	−0.37***
	m4	0.72 (0.70)	−2.12	2.59	0.39***	−0.35***	D6	0.39 (0.39)	−1.26	1.32	0.41***	−0.37***
	m5	0.64 (0.63)	−2.29	1.72	0.40***	−0.35***						
	m6	0.64 (0.60)	−1.85	1.69	0.40***	−0.36***						
	m7	0.72 (0.75)	−2.34	2.85	0.39***	−0.34***						
	m8	0.72 (0.71)	−2.11	2.60	0.39***	−0.35***						
Dark SC-IAT	m1	−0.06 (0.35)	−0.98	1.07	−0.22**	0.18*	D-Dark	−0.05 (0.31)	−0.74	0.78	−0.19*	0.17*
	m2	−0.06 (0.34)	−1.03	0.94	−0.23**	0.17*						
	m3	−0.06 (0.36)	−1.01	1.07	−0.22**	0.17*						
	m4	−0.06 (0.35)	−1.05	0.94	−0.22**	0.17*						
	m5	−0.06 (0.36)	−1.00	1.13	−0.20*	0.16*						
	m6	−0.06 (0.35)	−0.95	0.99	−0.20*	0.16						
	m7	−0.06 (0.36)	−1.04	1.13	−0.19*	0.16						
	m8	−0.06 (0.35)	−0.97	1.00	−0.20*	0.16						
Milk SC-IAT	m1	0.16 (0.39)	−1.92	1.22	0.17*	0.04	D-Milk	0.15 (0.33)	−0.93	1.21	0.13	0.06
	m2	0.16 (0.39)	−1.93	1.13	0.17*	0.04						
	m3	0.16 (0.41)	−1.92	1.5	0.17*	0.04						
	m4	0.16 (0.40)	−1.94	1.38	0.17*	0.04						
	m5	0.16 (0.38)	−1.39	1.23	0.15	0.05						
	m6	0.16 (0.37)	−1.40	1.14	0.15	0.05						
	m7	0.17 (0.39)	−1.39	1.51	0.16*	0.05						
	m8	0.16 (0.39)	−1.40	1.39	0.16*	0.05						

Note: *** $p < .001$, ** $p < .01$, * $p < .05$. *m*: Modified scoring algorithms; *D*: Typical scoring algorithms

Regardless of the scoring algorithm, the SC-IAT scores tended to have smaller effect sizes than the IAT scores. IAT modified scores showed higher effect sizes than IAT typical scores. Modified and typical SC-IAT scores were more consistent between each other.

Explicit Dark chocolate evaluation negatively and moderately correlated with explicit Milk chocolate evaluation ($r = -.39$, $p < .001$). IAT and Dark SC-IAT typical scores significantly correlated with both explicit chocolate evaluations. Milk SC-IAT typical score correlated with neither Dark nor Milk explicit evaluation. IAT modified scores significantly and moderately correlated with the explicit evaluations of both Dark and Milk chocolate. Modified scores of both SC-IATs significantly correlated with Milk chocolate explicit evaluation. Only the first four modified scores of Dark SC-IAT significantly correlated with Dark chocolate explicit evaluation. Correlation between Dark explicit evaluation and Milk SC-IAT scores (both typical and modified) was near zero.

Consistency between typical scores, modified scores, and explicit measures.   Correlation coefficients between typical IAT D scores ranged between .99 and 1.00 (all ps $< .001$). Correlations between the typical IAT scores and the Dark SC-IAT typical score were all $-.21$ (all ps $< .01$). No correlations were found between typical IAT D scores and typical Milk SC-IAT score (correlations ranged between $-.04$ and $-.03$, all ps $> .05$). Typical *D-Dark* and *D-Milk* positively correlated between each other ($r = .15$, $p > .05$), but the correlation was not significant. Correlations between IAT modified scores ranged between .97 and .99 (all ps $< .001$). Their correlations with modified *D-Dark* ranged between $-.31$ and $-.28$ (all ps $< .001$). Modified *D-Milk* and modified IAT D scores did not correlate with each other

(correlation coefficients ranged between $-.01$ and $.01$, all $ps > .050$). Correlations between modified *D-Dark* ranged between $.98$ and 1.00 (all $ps < .001$). Correlations between modified *D-Milk* ranged between $.99$ and 1.00 (all $ps < .001$). Correlations between modified *D-Milk* and *D-Dark* showed the same direction as the correlation between typical SC-IAT scores, ranging between $.15$ and $.20$. Interestingly, the correlation between all modified *D-Milk* and modified *D-Dark* from 5 to 8 (i.e., the scores in which error responses were replaced by the mean added with twice the standard deviation computed on all trials) showed slightly stronger and significant correlations, ranging from $.18$ and $.20$ (all $ps < .010$).

Behavioral outcome. Results of the logistic regressions for predicting the chocolate choice are reported in Table 1.4.

Table 1.4: Choice prediction results, single predictors..

	B (se)	Pseudo R^2	General	DCC	MCC		B (se)	Pseudo R^2	General	DCC	MCC
IAT *m1*	1.35*** [0.34]	0.16	0.64	0.65	0.63	*D3*	2.23*** [0.54]	0.18	0.64	0.63	0.66
IAT *m2*	1.38*** [0.35]	0.16	0.62	0.62	0.63	*D4*	2.26*** [0.55]	0.18	0.64	0.66	0.63
IAT *m3*	1.07*** [0.28]	0.15	0.63	0.65	0.62	*D5*	2.18*** [0.53]	0.18	0.63	0.63	0.63
IAT *m4*	1.12*** [0.29]	0.15	0.64	0.66	0.63	*D6*	2.22*** [0.55]	0.17	0.64	0.65	0.63
IAT *m5*	1.35*** [0.34]	0.16	0.64	0.65	0.63						
IAT *m6*	1.38*** [0.35]	0.16	0.64	0.63	0.64						
IAT *m7*	1.07 [0.28]	0.15	0.63	0.65	0.62						
IAT *m8*	1.12 [0.29]	0.15	0.64	0.66	0.63						
Dark *m1*	−0.73 [0.48]	0.02	0.53	0.62	0.44	*D-Dark*	−0.70*** [0.54]	0.01	0.53	0.65	0.41
Dark *m2*	−0.72 [0.49]	0.02	0.53	0.62	0.42						
Dark *m3*	−0.70 [0.47]	0.02	0.53	0.62	0.44						
Dark *m4*	−0.69 [0.48]	0.02	0.52	0.62	0.41						
Dark *m5*	−0.62 [0.47]	0.02	0.52	0.63	0.4						
Dark *m6*	−0.63 [0.48]	0.02	0.52	0.63	0.4						
Dark *m7*	−0.60 [0.46]	0.02	0.51	0.63	0.38						
Dark *m8*	−0.60 [0.47]	0.01	0.51	0.63	0.38						
Milk *m1*	0.33 [0.42]	0.01	0.53	0.77	0.26	*D-Milk*	0.35*** [0.49]	0.01	0.53	0.78	0.26
Milk *m2*	0.33 [0.43]	0.01	0.53	0.77	0.26						
Milk *m3*	0.32 [0.40]	0.01	0.52	0.76	0.26						
Milk *m4*	0.32 [0.41]	0.01	0.53	0.77	0.26						
Milk *m5*	0.31 [0.44]	0.00	0.5	0.75	0.23						
Milk *m6*	0.31 [0.44]	0.00	0.52	0.76	0.26						
Milk *m7*	0.30 [0.42]	0.00	0.5	0.75	0.23						
Milk *m8*	0.30 [0.42]	0.00	0.51	0.76	0.25						

Note: *** $p < .001$. Bs are the log-odds for the probability of choosing Milk chocolate; *m*: Modified scoring algorithms; *D*: Typical scoring algorithms, Gen: General accuracy of prediction, DCC: Dark Chocolate Choice accuracy of prediction, MCC: Milk Chocolate Choice accuracy of prediction.

The IAT scores outperformed both SC-IATs scores in predicting the chocolate choice. Models including IAT scores showed the highest values of Nagelkerke's R^2, and they resulted in a better accuracy of the prediction for both types of chocolate. Both SC-IATs scores showed low values of Nagelkerke's R^2, particularly Milk SC-IAT ones.

Typical and modified IAT scores tended to have similar values of both Nagelkerke's R^2 and accuracy of prediction. All modified Dark SC-IAT scores resulted in slightly higher values of Nagelkerke's R^2. Only the first four modified Milk chocolate SC-IAT scores showed slightly higher Nagelkerke's R^2 than the typical ones. Modified SC-IATs scores showed a slightly worse performance than the typical ones.

Results of choice prediction including the linear combination of SC-IAT scores are reported in Table 1.5.

Table 1.5: Choice prediction results: SC-IAT scores linear combination.

	B_{Dark} *(SE)*	B_{Milk} *(SE)*	*Pseudo R^2*	Gen	DCC	MCC
D-Dark + D-Milk	$-0.77\,(0.55)$	$0.46\,(0.50)$	0.02	0.55	0.67	0.41
$m1_{\text{Dark}}+m1_{\text{Milk}}$	$-0.81\,(0.49)$	$0.44\,(0.43)$	0.03	0.56	0.66	0.45
$m2_{\text{Dark}}+m2_{\text{Milk}}$	$-0.80\,(0.50)$	$0.44\,(0.44)$	0.03	0.53	0.66	0.40
$m3_{\text{Dark}}+m3_{\text{Milk}}$	$-0.78\,(0.48)$	$0.43\,(0.41)$	0.03	0.54	0.66	0.41
$m4_{\text{Dark}}+m4_{\text{Milk}}$	$-0.77\,(0.49)$	$0.43\,(0.42)$	0.03	0.54	0.67	0.40
$m5_{\text{Dark}}+m5_{\text{Milk}}$	$-0.71\,(0.48)$	$0.44\,(0.45)$	0.02	0.55	0.68	0.40
$m6_{\text{Dark}}+m6_{\text{Milk}}$	$-0.72\,(0.49)$	$0.44\,(0.45)$	0.02	0.54	0.68	0.38
$m7_{\text{Dark}}+m7_{\text{Milk}}$	$-0.68\,(0.47)$	$0.42\,(0.43)$	0.02	0.55	0.68	0.40
$m8_{\text{Dark}}+m8_{\text{Milk}}$	$-0.69\,(0.48)$	$0.42\,(0.44)$	0.02	0.54	0.68	0.38

Note: Bs are the log-odds for the probability of choosing Milk chocolate. *m*: Modified scoring algorithms; *D*: Typical scoring algorithms; Gen: General accuracy of prediction, Pseudo *Pseudo R^2*: Nagelkerke's R; DCC: Dark Chocolate Choice accuracy of prediction; MCC: Milk Chocolate Choice accuracy of prediction.

The linear combination of SC-IATs scores resulted in a better prediction of the chocolate choice than that provided by their singular scores, despite their performance was still outperformed by that of the IAT. The coefficients of typical and modified *D-Dark* scores tended to be higher than the coefficients of typical and modified *D-Milk* scores. The linear combination of the first four scores resulted in higher Nagelkerke's R^2 values, both in comparison with the typical scores and with the last four modified scores. General accuracy and DCC accuracy were similar across all scores, while MCC accuracy showed a higher variability. Specifically, the linear combination of *m6* and *m8* showed the worst performance of all. The linear combination of *m1* resulted in the highest MCC accuracy.

As a final analysis, the incremental validity of the IAT and the two SC-IATs in respect to the self-report chocolate evaluations was investigated. Four hierarchical multiple logistic regressions for predicting the chocolate choice were specified for each of the scoring procedure. In the first step, Dark and Milk chocolate explicit evaluations were included. IAT *D* scores entered at the second step. *D-Dark* entered at the third step, and *D-Milk* entered at the fourth step. This procedure was followed for both typical and modified scores. Nagelkerke's R^2 was used as a criterion to decide whether the added predictor was useful to account for the chocolate choice. Nagelkerke's R^2 at the first step (i.e., the model including only the explicit chocolate evaluations) was 0.83. From the second step on, the Nagelkerke's R^2 remains 0.84 for both typical and modified scores. It is reasonable to argue that implicit measures scores do not add anything to the prediction given by the explicit measures. However, this result should be interpreted with caution because the explicit chocolate evaluation was asked right before the behavioral choice.

1.3.4 Final remarks

By aligning the administration of the IAT and the SC-IAT, as well as their scoring methods, as much as possible it was possible to fairly investigate their relationship with explicit measures and their ability to predict behavioral outcomes. Consistently with the assumptions underlying the functioning of the two measures, IAT scores highly correlated with both explicit chocolate evaluations, while scores of the SC-IAT tended to correlate with just one of the explicit chocolate evaluation.

The IAT outperformed both SC-IATs in the prediction of the behavioral choice. This result held both when the individual scores of the two SC-IATs were used and when their linear combination was considered. Taking these considerations together, it is possible to argue that the IAT has a better predictive capacity than the SC-IAT.

However, the higher predictive ability of IAT scores might be due to the characteristics of the choice task itself. Since participants were presented with two different bowls of chocolate and were invited to take just one chocolate bar, their liking and/or dislike for both types of chocolate were concurrently playing a role in determining the choice. A measure able to include the comparative evaluation of the chocolate types, like the IAT, might hence best account for the actual chocolate preference from which it derives a better predictive ability. Conversely, measures dealing with only one of the components of the chocolate evaluation, like the SC-IAT, might be disadvantaged. However, even when SC-IAT scores were considered concurrently the general accuracy of prediction was similar to the one obtained when the single scores were considered. Nonetheless, it resulted in a slightly better accuracy of pre-

diction of Milk chocolate choice, especially when compared with the performance of *D-Milk*

scores alone. This result is supporting the claim according to which a measure accounting for

the relative attitudes towards two contrasting objects results in better prediction of the choice

between alternative options.

Since the better performance of the IAT might have been due to both the choice task and

the type of preference assessed, it would be interesting to compare the performance of the two

implicit measures when a clearly contrasted category is not identifiable, such as self-esteem.

In such cases, the absolute measure provided by the SC-IAT should outperform the relative

one provided by the IAT in predicting behavioral outcomes. Additionally, the SC-IAT might

outperform the IAT in predicting behavioral outcomes when the choice task is not strictly

dichotomous. For instance, respondents might be left free to choose between two chocolate

bars, both of them, or none of them, or even between different types of candy bars, including

Dark and Milk chocolate bars.

Both in ? (?) and in this study, the predictive validity of implicit measures was assessed

for non-socially relevant stimuli, like soda and chocolate preference. Future studies should

investigate the IAT and SC-IAT predictive validity in respect to socially relevant stimuli, such

as members of stigmatized social groups. In pursuing this aim, different behavioral indicators

might be used as a dependent variable, such as the willingness to affiliate with members of

the stigmatized social group. Another one could be having or not contacts with members of

the stigmatized social groups.

Finally, ? (?) administered the SC-IAT with a rtw at 1,500 ms, while in this study it was

not used. We could have applied an a posterior threshold for upper tail responses as if a rtw

had been used in the administration. However, we decided not to do so because using a rtw

affects respondents' performance. Indeed, ? (?) observed that the response time window

produced a sense of urgency for giving the response that was missing when the rtw was not

included. Future research on the systematic comparison between the two implicit measures

might include a response time window for both the IAT and the SC-IAT.

1.4   R development

1.4.1   DscoreApp

Different options are available for computing the IAT D score, including SPSS syntax, Inquisit scripts, and R packages.

Inquisit scripts are the most straightforward way for obtaining the D score, since they compute it right after the IAT administration procedure, store the results along with other information on participants' performance (e.g., response time for each IAT trial, correct and incorrect responses), and do not require any programming skills. Nonetheless, these scripts work only when associated with the Inquisit administration procedure, can compute just one of the available D score algorithms at the time, and do not provide functions for visually inspecting the results. Finally, Inquisit requires a license to be used.

SPSS syntaxes provides several information on respondents' performance, and they are not tied to a specific administration software. Nonetheless, their use requires the SPSS license and a certain degree of expertise with SPSS language. They do provide the chance to compute different D score algorithms by providing different scripts for the computation of each algorithm. No functions for directly plotting the results are provided.

R packages (illustrated in Table 1.6) provides the open source alternative to both Inquisit and SPSS syntaxes. However, their use is not always straightforward because they require quite advanced programming skills, their functions are limited and they are often not well defined and understandable.

Table 1.6: Overview of the available packages for computing the IAT *D* score.

Package	Author	Functions	Multiple *D* score	Plot	Reliability
`IATanalytics`	? (?)	`IATanalytics`: Function to analyze raw data from an IAT `sampledata`: Sample dataset from a typical IAT	No	No	No
`IATScore`	? (?)	`BriefIAT`: Sample Brief IAT Data set (Abbreviated IAT) `IAT`: Sample IAT Dataset (Typical) `IATScore`: Score Implicit Association Test (IAT) output `TooFastIAT`: Sample IAT Dataset (Participant went too fast)	No	No	No
`IAT`	? (?)	`cleanIAT`: Clean IAT data using the updated D-Scoring algorithm	Yes	Yes	Yes

		`IATData`: Sample Gender Stereotype Implicit Association Test data			
		`plotIIV`: Plot intraindividual variability of reaction time			
		`plotIndVar`: Plot individual variability in the IAT			
		`plotItemErr`: Plot proportion of errors per item in the IAT			
		`plotItemVar`: Plot IAT item variability			
`IATScores`	? (?)	`alg2param`: Convert the algorithm names to the generating parameters	Yes	Yes	Yes
		`Pretreatment`: Pretreat the IAT data in input			
		`RobustScores`: Compute the Robust IAT scores			
		`SplitHalf`: Split half reliability			
		`TestRetest`: TestRetest			
		`Tgraph`: Layout `qgraph` for multiple comparisons by package `nparcomp`			

Packages `IATanalytics` and `IATscore` require for a specific arrangement of the data set columns to compute the *D* score, and they both include only the functions for computing the *D* score. The package `IATscore` includes a also a function for scoring the Brief IAT (B-IAT; ?, ?). Both packages compute the *D* score for one respondent at a time, and no details on the specific algorithm that is computed are provided. The package `IAT` includes functions for cleaning the original data set, for plotting the data, and for computing different *D* score algorithms. However, it does not provide a clear labeling of each of the algorithms, but it allows the users to specify whether to discard the trials under 400 ms or the error penalty to use. Additionally, the package `IAT` asks for a counter-intuitive coding of accuracy responses (i.e., 0 for correct responses, 1 for error responses). The plotting functions included in this package are not meant for plotting the *D* score results but the raw data.

The package `IATScores` appears to be the most complete package. Besides the functions for cleaning the data and for computing the typical IAT *D* score algorithms, it includes functions for the computation of the robust *D* score algorithms presented in (?, ?). Additionally, it includes functions for computing the IAT reliability (i.e., test-retest reliability and split half). As for package `IAT`, the plotting functions in package `IATScores` are not meant for plotting the results of the *D* score computation.

DscoreApp (?, ?) was developed in `R` by means of packages `shiny` (?, ?) and `shinyjs` (?, ?) with the aim of providing an open source tool able to make the *D* score computation easier for researchers who commonly employ the IAT but have little or no programming experience. Furthermore, by providing an immediate representation of the results, it allows for a glimpse of the IAT results. DscoreApp can be retrieved at `http://fisppa.psy.unipd.it/DscoreApp`

The source code of DscoreApp is available on GitHub (`https://github.com/OttaviaE/DscoreApp`).

DscoreApp has been published in ? (?).

DscoreApp is organized in different panels ("Input", "Read Me First", "D-score results", and "Descriptive statistics"). Setting options and functions in "Input" panel, as well as the menu in "Read Me First" panel, are interactive, so that users can easily access the information on DscoreApp functions and amenities.

"Read Me First" panel provides important information on DscoreApp functioning, including an overview of *D* score algorithms. A downloadable template suggested for using the app is provided (i.e., **Download template** button), even though it is not necessary to use it. DscoreApp is designed to work as long as the uploaded data set is in a CSV format, and includes the following variables: `participant` (i.e., participants' IDs), `latency` (i.e., response latencies in milliseconds), `correct` (response accuracy, either 0 for error responses or 1 for correct responses), `block` (i.e., labels identifying the four associative blocks of the IAT, B3, B4, and B6, B7). This panel also contains information on the downloadable file containing the results of the *D* score computation.

Users can either upload their own data set (i.e., by using **Browse** function), or use the toy data set included in DscoreApp (i.e., by checking `Race IAT dataset` checkbox) in "Input" panel. Once the data set is read, the app automatically populates the drop-down menus for choosing the labels denoting the four associative blocks, and the **Prepare data** button becomes clickable. When data are ready for *D* score computation, the message "Data are ready" appears next to **Prepare data** button, and all options for its computation and graphical display become active. Once a *D* score algorithm is chosen from the drop-down menu

"Select your D", the **Calculate & Update** button becomes active. Users can decide whether to eliminate participants whose error percentage exceeds a specified threshold (default is 25% according to ?, ?) or whose fast responses (< 300 ms) exceed 10% of the total responses (?, ?). When these options are selected, participants exceeding the thresholds (if any) are not displayed in "D-score results" panel. Every time a change in the configuration is made, the **Compute & Update** button must be clicked to apply the changes.

"D-score results" panel (Figure 1.1) is populated once **Calculate & Update** button is clicked for the first time. Both descriptive statistics of the results and their graphical repre-
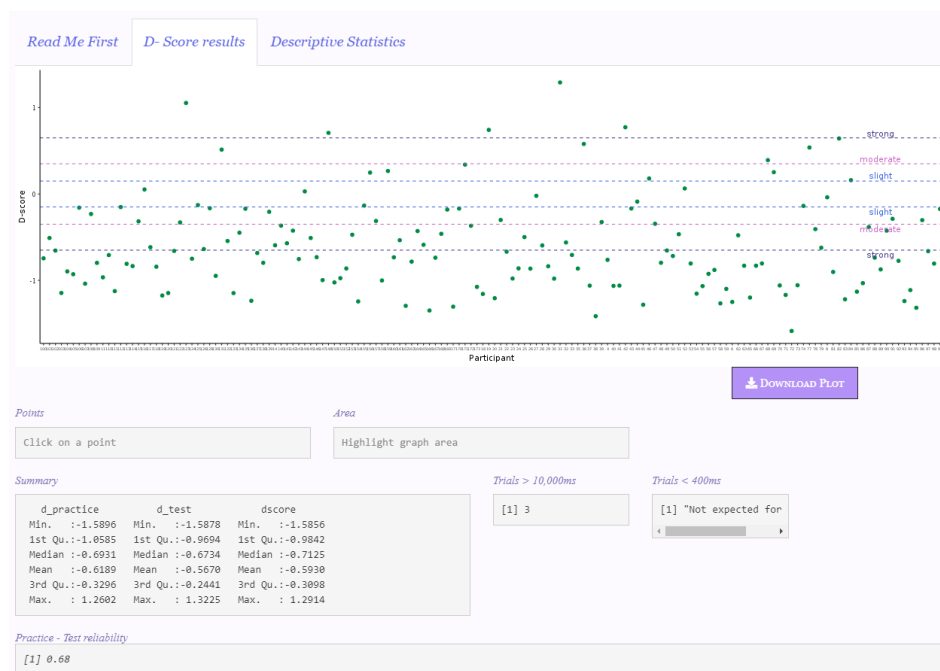

Figure 1.1: DscoreApp results panel.

sentation are available at the same time, and they change interactively as users change configuration in "Input" panel. Summary box reports the descriptive statistics of the D_{practice}, D_{test}, and the actual D score. Trials $> 10,000$ms box reports the number of trials dis-

carded because of a slow latency (if any), while `Trials < 400ms` box reports the number of trials discarded because of fast response times. This box is populated only when a *D* score with fast trials deletion is selected, otherwise the "Not expected for this D" label is displayed. `Practice-Test reliability` box contains the IAT reliability computed as the correlation between associative practice and associative test blocks across participants (?, ?).

Graphical representation is a convenient way to identify extreme scores or particular response patterns. Since it might be difficult to link a particular point (or points area) in the graph with the corresponding participants' IDs in the data set, DscoreApp comes with two handy tools designed to access the respondents' IDs from the graph. By clicking on a point in the graph, the ID of the participant that corresponds to the selected point, and his/her *D* score, appears in `Points` box. By highlighting an area of the graph, the IDs of participants included in the area, along with their *D* scores appears in `Area` box. `Points` box and *Area* box are represented in Figure 1.1, right underneath the graphical representation of the results.

DscoreApp provides users with different options for the graphical representation of the results (Figure 1.2), at both individual (Figure 1.2a) and sample (Figure 1.2b, 1.2c, 1.2d) levels.

All the graphical representations are downloadable in a .pdf format.

1.4.2 `implicitMeasures`

Despite both the IAT and the SC-IAT are commonly used for the implicit assessment of several constructs, only `R` packages for the computation of the IAT *D* score are available,

(a) Points graph.　　　　　　　　(b) Histogram.

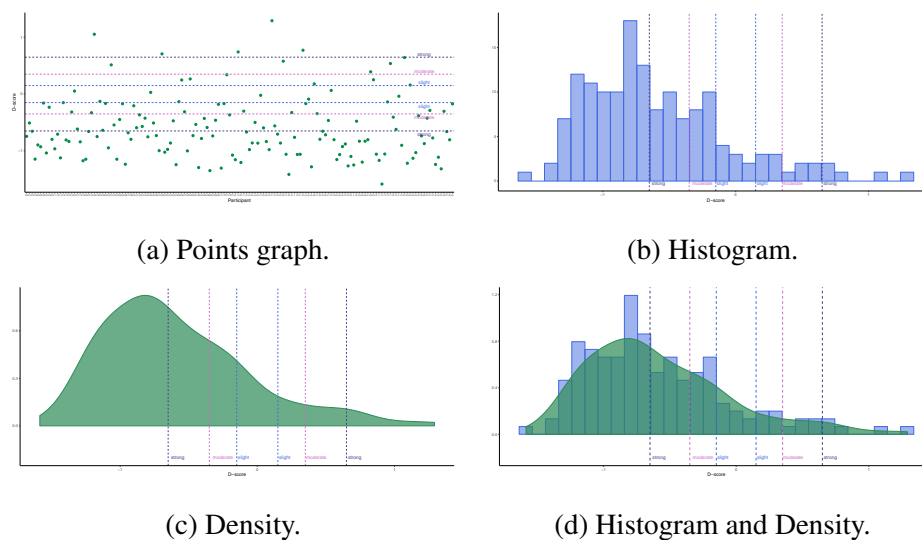(c) Density.　　　　　　　　(d) Histogram and Density.

Figure 1.2: Available graph representations.

while there are no packages for computing the SC-IAT *D* score.

Besides the above-mentioned shortcomings of R packages for computing IAT *D* score, there are also issues related to the replicability of the results. Researchers' choice on which IAT *D* score algorithm to compute might influence the results and the conclusions that are drawn from them (?, ?). Moreover, in many cases researchers fail to report of the exact *D* score algorithm they decided to use, raising replicability issues (?, ?). Replicability issues might also rise from the many steps needed to clean and prepare the starting data set for computing the *D* score (?, ?).

Despite DscoreApp addresses the majority of the replicability issues, it presents some drawbacks as well. Firstly, since the code is into the shiny interface, it cannot be called from the command line, making impossible to replicate it and hence the results. To be fair, this is a general and outstanding issue concerning shiny apps in general. This might not be a concern for general users, but it is indeed a problem in an open science framework where every code

should be accessible and replicable with no effort. Moreover, the downloadable graphical

representations are provided in a pdf format, and hence they cannot be further modified.

The package `implicitMeasures` (?, ?, ?) was aimed at addressing the issues con-

cerning both `R` packages and DscoreApp. It provides an easy and open source way to clean

and score both the IAT and the SC-IAT, to easily compare different IAT *D* score algorithms,

and to provide clear and customizable plots. Plot functions are all based on `ggplot2` (?, ?).

The package `implicitMeasures` has been published in ? (?).

The source code of `implicitMeasures` is available on GitHub

(`https://github.com/OttaviaE/implicitMeasures`). The package is down-

loadable from CRAN

(`https://cran.r-project.org/web/packages/implicitMeasures/index.html`).

Table 1.7 provides an overview of the functions included in the `implicitMeasures`

package.

Table 1.7: Contents and functions `implicitMeasures`.

Function	Description
`clean_iat()`	Prepare and clean IAT data
`clean_sciat()`	Prepare and clean SCIAT data
`compute_iat()`	Compute IAT *D* score
`compute_sciat()`	Compute SCIAT *D* score
`descript_d()`	Print descriptive table of *D* scores (also in LaTeX)
`d_density()`	Plot either IAT or SCIAT scores (distribution)
`d_point()`	Plot either IAT or SCIAT scores (points)
`IAT_rel()`	Compute IAT reliability
`multi_dsciat()`	Plot scores resulting from two SCIATs
`multi_dscore()`	Compute and plot multple IAT *D* scores
`raw_data()`	Example data set

`implicitMeaures` provides an easy way to compute the algorithms for the both the IAT and the SC-IAT in an automated way. By explicitly referring to the *D* score algorithm that has been used for computing the IAT score in the package, other users can easily replicate the results. Additionally, the possibility to compute all available algorithms for the IAT *D* score allows for an easy comparison between them. This makes possible to investigate whether or how the elimination of fast responses or the error replacement strategies affect the results.

All objects created with functions in `implicitMeasures` functions can be exported in external files. For example, the data frame obtained from `clean_iat()` function can be easily exported in a CSV file and then uploaded to DscoreApp (see Section 1.4.1).

The functions for plotting the results are based on `ggplot2` (?, ?), and they can be further modified by users, for instance by taking out the legend, adjusting the figure margins, changing labels and font. All the plots are then exportable as images (.jpg or .png) or as a .pdf.