



# Lässt sich die Testschwierigkeit und somit die prädiktive Power des impliziten Assoziationstests über die Target-Stimuli beeinflussen?



Barbara Gräser, Magdalena Klodt, Conrad Opitz, Annika Thiel  
Leitung: M. Sc. Merlin Urban

## THEORIE

Jüngere Forschungen zweifeln den Mehrwert der Ergebnisse des IATs gegenüber direkten Maßen an. Ist der IAT generell ein Maß geringer prädiktiver Power, oder gibt es Moderatoren - wie in diesem Fall die Testschwierigkeit - die die prädiktive Power erhöhen können?



Urban et al. konnten zeigen, dass eine mittlere IAT Testschwierigkeit (definiert als IAT Effekt von null) zu mehr Varianz und somit zu einer besseren prädiktiven Power führt als eine extreme Testschwierigkeit (definiert als ein IAT-Effekt, der stark von null abweicht).

Govan und Williams zeigten, dass der mittlere IAT-Effekt (i.e. Testschwierigkeit) über die Valenz der Target-Stimuli beeinflusst werden kann.

Bei einer positiven und einer negativen Target-Kategorie:

- ist der IAT-Effekt stärker von null abweichend, wenn die Target-Stimuli in ihrer Valenz typisch sind, also der Valenz der Target-Kategorien entsprechen.
- ist der IAT-Effekt weniger stark von null abweichend, wenn die Target-Stimuli in ihrer Valenz atypisch sind, also der Valenz der Target-Kategorien nicht entsprechen.

### Hypothese 1:

Die IAT-Testschwierigkeit lässt sich über die Manipulation der Valenz der Target-Stimuli auf die beschriebene Weise beeinflussen.

### Hypothese 2:

Je extremer die IAT-Testschwierigkeit, desto geringer die IAT-Varianz.

### Hypothese 3:

Je extremer die IAT-Testschwierigkeit und je geringer die IAT-Varianz, desto geringer die prädiktive Power des IATs.

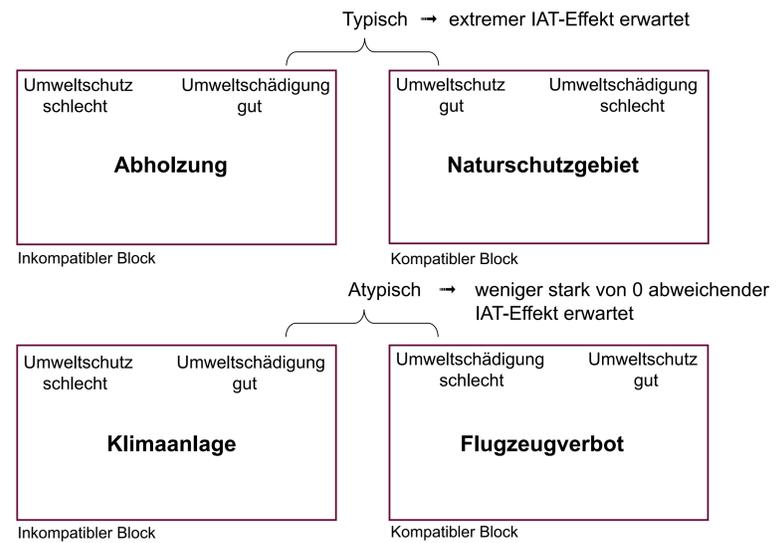


Abb. 1

$$\text{IAT-Effekt} = RT_{\text{inkompatibel}} - RT_{\text{kompatibel}}$$

## METHODEN

### Stichprobe

$N_{\text{final}} = 86$   $M_{\text{Alter}} = 21.76$   $SD_{\text{Alter}} = 2.73$   
95,35 % Psychologiestudierende

### 2x2 mixed Design

Within-Faktor **Stimulusset**:  
Stimulusset mit *atypisch* valenten Stimuli vs.  
Stimulusset mit *typisch* valenten Stimuli

Between-Faktor **Blockreihenfolge**:  
Beginn mit *kompatiblen* vs.  
*inkompatiblen* Block

→ Zuordnung von den  
Bedingungen wurde randomisiert



### Messinstrumente

**IAT** → IAT-Effekte berechnet mit D-Score Algorithmus (nur von den Target-Kategorien)  
Atypisches Stimulusset:  $\alpha = .46$ , Typisches Stimulusset:  $\alpha = .45$

#### Target-Kategorien

##### Umweltschutz

→ 3 Stimuli (typisch) *Solarenergie, Naturschutzgebiet, Fairtrade*

→ 3 Stimuli (atypisch) *Flugzeugverbot, Radikale Umweltaktivisten, Konsumverzicht*

##### Umweltschädigung

→ 3 Stimuli (typisch) *Vermüllung, Abholzung, Mikroplastik*

→ 3 Stimuli (atypisch) *Klimaanlage, Wirtschaftswachstum, Luxuriöser Lebensstil*

#### Attribut-Kategorien

**gut** → 6 Stimuli *angenehm, toll, fabelhaft, positiv, glücklich, freundlich*

**schlecht** → 6 Stimuli *hässlich, negativ, mies, fehlerhaft, furchtbar, schrecklich*

Tabelle 1: Beispielhafte Darstellung des Ablaufs (Beginn mit kompatibellem Block)

Block	Trials	Funktion	Items (linke Taste D)	Items (rechte Taste L)
1	12	Übung	Umweltschädigung	Umweltschutz
2	12	Übung	schlecht	gut
3, 6, 8, 10, 12	24	Test	Umweltschädigung + schlecht	Umweltschutz + gut
4	24	Übung	Umweltschutz	Umweltschädigung
5, 7, 9, 11, 13	24	Test	Umweltschutz + schlecht	Umweltschädigung + gut

### Kriteriumsmaße

Bauch- und tatsächliche Gefühle gegenüber Umweltschutz/ Umweltschädigung  
1 = extrem negativ bis 10 = extrem positiv  
→  $\alpha = .57$  (Gefühlsmaß)

Bewertungen der Valenz aller Target-Stimuli  
1 = extrem negativ bis 9 = extrem positiv  
→  $\alpha = .77$  (Valenzmaß)

Verhalten (Häufigkeit von umweltschützenden oder -schädigendem Verhalten)  
1 = nie bis 7 = häufig  
→  $\alpha = .22$  (Verhaltensmaß)

**Hypothese 1:** Das atypische Stimulusset hatte wie erwartet eine weniger extreme Testschwierigkeit als das typische Stimulusset

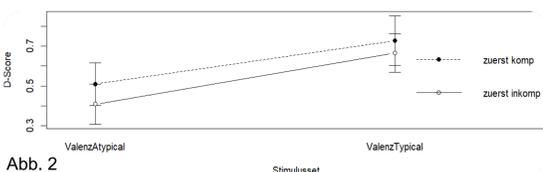


Abb. 2

**Hypothese 2:** Die IAT-Varianz des atypischen Stimulussets war nicht signifikant höher als die des typischen

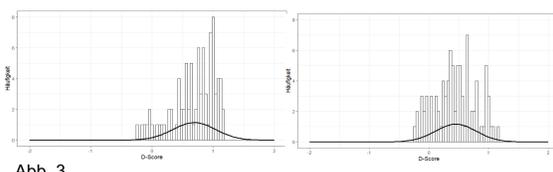


Abb. 3

**Hypothese 3:** Die prädiktive Power war im atypischen Stimulusset nicht höher als im typischen.

Tabelle 2 Korrelationen der Stimulussets mit Kriteriumsmaßen

	Stimulusvalenzmaß	Gefühlsmaß	Verhaltensmaß
Atypisch	.03	.07	-.05
Typisch	.02	-.11	.01

Anmerkung \* $p < .05$

Zweifaktorielle mixed ANOVA:

**Haupteffekt Stimulusset wie angenommen signifikant**

→  $F(1, 84) = 49.49, p < .0001, \eta^2_{\text{part}} = .37$

**Haupteffekt Blockreihenfolge wie angenommen nicht signifikant**

→  $F(1, 84) = 1.44, p = .23, \eta^2_{\text{part}} = .02$

**Interaktion Faktoren wie angenommen nicht signifikant**

→  $F(1, 84) = 0.29, p = .59, \eta^2_{\text{part}} = .00$

IAT-Varianz atypisches Stimulusset ( $\hat{\sigma} = 0.118$ ) anders als in H2 angenommen **nicht signifikant** höher als typisches Stimulusset ( $\hat{\sigma} = 0.123$ )

→ Levene-Test:  $F(1, 170) = .07, p = .80$

Die Korrelationen des atypischen und typischen Stimulussets mit den Kriteriumsmaßen zeigten sich entgegen der 3. Hypothese **nicht signifikant** von Null verschieden (s. Tabelle 2)

Die Korrelationen des typischen und atypischen Stimulussets mit dem Stimulusvalenzmaß (Z-Test:  $t = 0.08, p = .47$ ) und dem Verhaltensmaß (Z-Test:  $t = 0.65, p = .26$ ) zeigten sich entgegen der 3. Hypothese **nicht signifikant** verschieden

Die Korrelationen des typischen und atypischen Stimulussets mit dem Gefühlsmaß (Z-Test:  $t = 1.86, p = .03$ ) zeigten sich wie erwartet **signifikant** verschieden

## DISKUSSION

### Hypothese 1



Testschwierigkeit ließ sich insofern verändern, als dass das atypische Stimulusset einen geringeren mittleren IAT-Effekt hatte als das typische Stimulusset und eine weniger extreme IAT-Testschwierigkeit aufwies

### Hypothese 2



Das atypische Set hatte nicht mehr Varianz als das typische Kontinuierliche Skala der D-Scores, bei denen extreme Mittelwerte auf der Skala nicht mit einer schiefen Verteilung einhergehen müssen → Verteilung kann sich lediglich verschieben  
Bei Fragebögen mit diskreten Skalen und eingeschränktem Antwortbereich hingegen, muss eine schiefe Verteilung bei extremen Mittelwerten mit einhergehen

### Hypothese 3



Wir konnten keine erhöhte prädiktive Power für das atypische Stimulusset im Vergleich zum typischen feststellen. Bzgl. Gefühlsmaß: signifikante Unterschiede in die angenommene Richtung, Korrelation des atypischen Sets ist aber nicht signifikant von 0 verschieden und prädiktive Validität somit weiterhin eingeschränkt.  
→ lässt sich durch die fehlende erhöhte Varianz erklären (siehe H2)

Korrelationen waren generell sehr niedrig: möglicherweise auf niedrige IAT-Reliabilitäten oder die Homogenität der Stichprobe und damit verbundenen Varianzeinschränkungen in allen Maßen zurückzuführen



## LITERATUR

Govan, C. L. & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal Of Experimental Social Psychology*, 40(3), 357-365. <https://doi.org/10.1016/j.jesp.2003.07.002>

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. <https://doi.org/10.1037/0022-3514.74.6.1464>

Urban, M., Koch T., & Rothermund K. (in press). The Implicit Association Test and its Difficulty(ies): Introducing the Test Difficulty Concept to Increase the True-Score Variance and, Consequently, the Predictive Power of Implicit Association Tests. *Journal of Personality and Social Psychology*.



FRIEDRICH-SCHILLER-UNIVERSITÄT JENA