



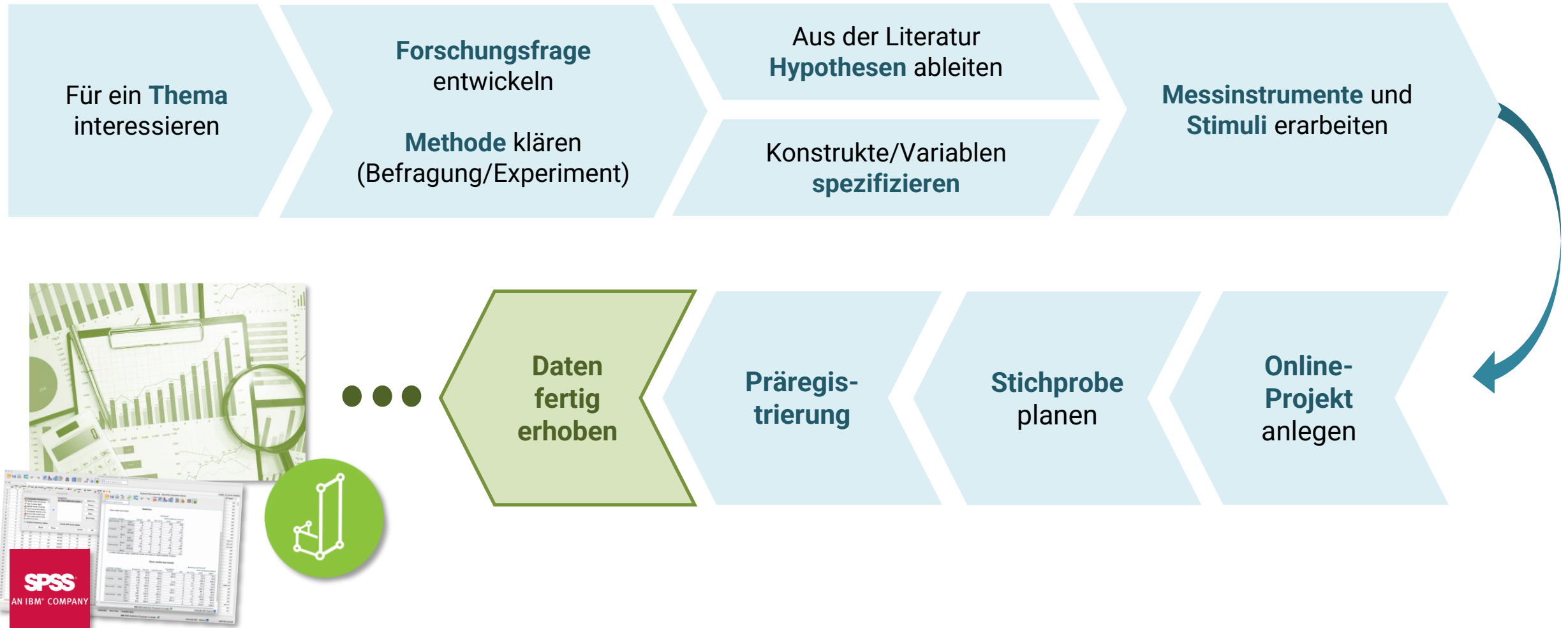
„...und jetzt?“ – Die erste **Datenaufbereitung**

Video-Tutorial

von Prof. Dr. Jan-Philipp Stein



Einführung > Fast am Ziel des Forschungsprozesses





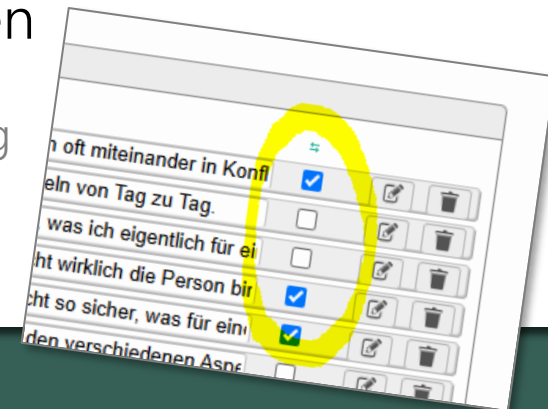
Schritt 1 > Die Daten herunterladen

- SoSciSurvey erlaubt den **unmittelbaren Download** als **SPSS-Syntaxdatei (.sps)**
 - Öffnen Sie diesen Download mit SPSS, markieren Sie den kompletten Text und führen Sie das Skript aus (grüner Pfeil)
 - Nun wird Ihnen die eigentliche **Datenmappe** (.sav) erstellt, welche Sie bitte gesondert abspeichern
 - **Tipp:** Mehrere Versionen mit dem Datum der Bearbeitung archivieren (z. B. „Datenmappe Bachelorarbeit 2025_01_15“)
- Wenn Sie mit einer anderen Software (z. B. *JASP*) arbeiten möchten, können Sie die Daten bei SoSciSurvey auch in anderen Formaten wie **.csv** erhalten
 - Viele Statistikprogramme erlauben aber auch den Import von SPSS-Files



Schritt 2 > Sich einen Überblick verschaffen – und „aufräumen“

- Der erste Datensatz erscheint noch ziemlich unaufgeräumt...
- Daher **zentrale erste Schritte**:
 - **Überblick verschaffen**: Welche Variablen gehören zu welchem Fragebogen? Was hat SoSciSurvey alles miterfasst (z. B. TIME-Variablen am Ende)?
 - **Variablennamen und -beschriftungen („Labels“) anpassen**: Jeder Variablenname sollte selbsterklärend sein
 - Beispiele: „SE01“ zu „selfesteem_item01“ – oder – „SD02“ zu „gender“ (*natürlich individuell prüfen!*)
 - **Kontrolle benötigter Umpolungen**: Wenn im gleichen Fragebogen manche Items positiv und andere negativ formuliert sind, müssen letztere umcodiert werden
 - *außer* dies wurde bereits direkt in SoSciSurvey für die betreffenden Items eingestellt
 - dann ist die erfolgreiche Umpolung am Wort „(umgepolt)“ in der Variablenbeschriftung zu erkennen





Exkurs > Korrekte Umcodierung einzelner Items

- Zur **Umcodierung** wählen Sie im SPSS das Menü „Transformieren > Umcodieren in andere Variablen“
 - Sinnvolle Benennung der neuen Variable! → z. B. „selfesteem_item01_R“ (R = recoded)
- Beachten Sie, dass für **alle** möglichen Skalenwerte ein neuer Wert vergeben werden muss; Beispiel 5-stufige Skala:
 - Alter Wert **1** → Neuer Wert **5**
 - Alter Wert **2** → Neuer Wert **4**
 - Alter Wert **3** → Neuer Wert **3** (*theoretisch unnötig, aber sicherheitshalber mitnehmen.*)
 - Alter Wert **4** → Neuer Wert **2**
 - Alter Wert **5** → Neuer Wert **1**





Schritt 3 > Die erste **Filterung** ungültiger Datensätze


- **Erste Sichtung:** Wer hat die Befragung vorzeitig abgebrochen?
 - Diese Datensätze werden aus dem Rohdatensatz **spurlos entfernt** (ist auch ethisch geboten!)
- **Zweite Sichtung:** Wer hat unsere inhaltlichen (präregistrierten) Einschlusskriterien nicht erfüllt?
 - Diese Datensätze werden **nicht** aus dem Rohdatensatz **entfernt** – sondern mit einer sogenannten **Filtervariable** inaktiv gestellt (siehe nächste Folie!)
 - **Vorteil:** Ermöglicht später ggf. Reanalysen mit strengeren oder lockereren Einschlusskriterien





Schritt 3 > Die erste **Filterung** ungültiger Datensätze

- Einfügen einer **numerischen Variable „filter“** ganz vorne in der Datenmappe
 - Alle Versuchspersonen erhalten hier zunächst eine 1 (= *gültige Teilnahme*)
 - Anschließend alle Personen nach den präregistrierten Ausschlusskriterien prüfen
→ wenn ein Kriterium zutrifft, wird bei der Person die 1 in eine 0 geändert (= *ungültig*)
 - **Bonustipp:** Erstellen Sie eine zusätzliche Textvariable „filter_grund“ (vom Typ Zeichenfolge), in der Sie den Ausschlussgrund für sich dokumentieren
- **Achtung:** Zukünftig muss bei jedem neuen Start von SPSS die vorbereitete Filterung erst aktiviert werden (Menü „Daten → Fälle auswählen → Filtervariable verwenden“)
 - Die erfolgreiche Filterung erkennen Sie daran, dass einzelne Fälle durchgestrichen sind

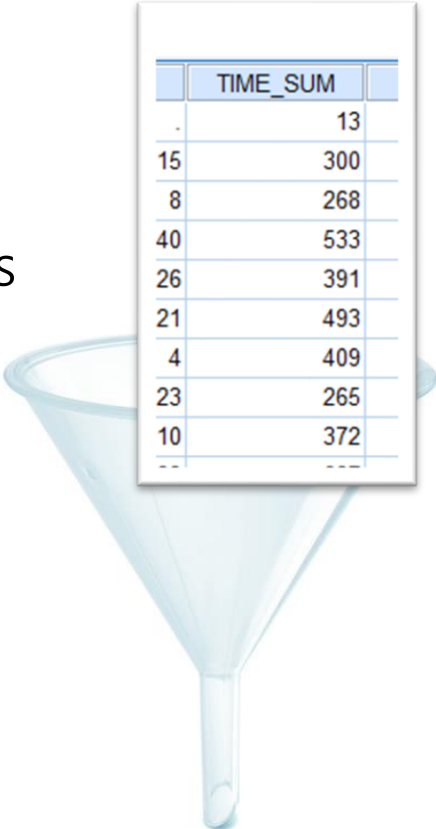


	filter	filter_grund	CASE
1	0	nicht gewissenhaft	253
2	1		93
3	1		94
4	1		96
5	1		97
6	1		98



Schritt 3 > Die erste **Filterung** ungültiger Datensätze

- Bezüglich der **Bearbeitungszeit** vereinbaren wir je nach Projekt unterschiedliche Ausschlusskriterien
 - Relevant ist hier auf jeden Fall die Variable „**TIME_SUM**“, die SoSciSurvey standardmäßig erfasst
 - Besprechen Sie mit Ihren Betreuer:innen, wie Sie hier über Ein-/Ausschluss entscheiden sollen (gängig: Betrachtung von Histogrammen und Boxplots)
 - Ähnliches gilt für den Ausschluss anderweitig unplausibler Datensätze
 - ✓ z. B. komische Antworten auf offene Fragen, verdächtige Antwortmuster



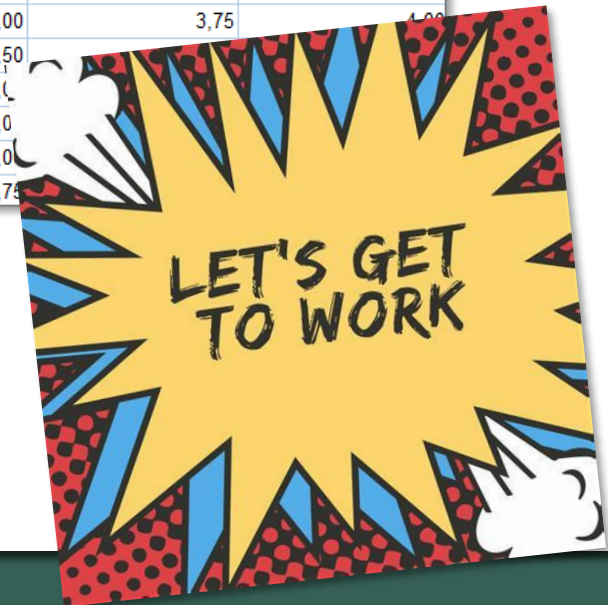
	TIME_SUM
.	13
15	300
8	268
40	533
26	391
21	493
4	409
23	265
10	372
--	---



Schritt 4 ➤ Die ersten Rechenoperationen mit Ihren Daten



filter	filter_grund	gender	age	political_orientation	selfesteem_mean	bodyesteem_mean	Big5_agreeable_mean	Big5_extraversion_mean
0	attention check	Male	18	Extremely Right	3,00	2,50	3,25	3,25
0	manipulation check	Other	19	Extremely Left	2,83	2,75	3,00	2,75
1		Male	22	Extremely Right	4,75	5,00	4,75	4,50
1		Male	22	Left	4,67	4,50	5,00	4,50
1		Male	22	Right	4,58	4,50	5,00	4,25
1		Male	22	Center	3,25	3,50	3,25	3,00
1		Male	23	Extremely Right	2,33	3,00	2,00	2,00
1		Male	23	Center	4,67	4,50	4,75	4,75
1		Female	23	Extremely Right	3,00	2,75	3,00	3,25
1		Male	24	Left	3,92	4,00	3,75	4,00
1		Male	24	Center	3,92	4,50	4,00	4,00
1		Male	24	Left	3,17	3,00	3,00	3,00
13		Female	24	Left	4,00	4,00	4,00	4,00
14		Male	25	Extremely Left	4,58	5,00	4,00	4,00
15		Male	25	Right	2,58	3,75	3,00	3,00





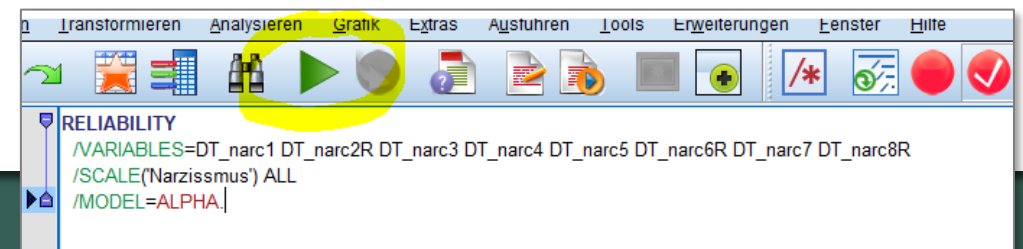
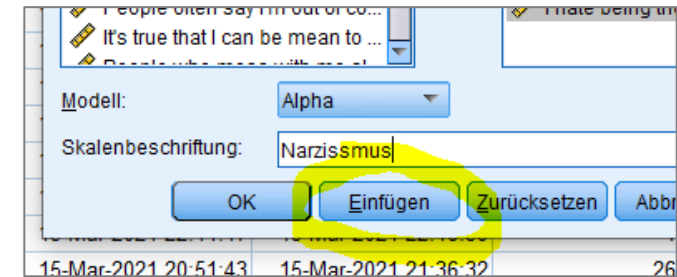
Schritt 4 > Die ersten Rechenoperationen mit Ihren Daten

- Erste **niedrigschwellige Rechenoperationen** und **deskriptive Analysen**
 - **Bildung der Indizes** für die jeweils mit mehreren Items gemessenen Variablen
 - ✓ Tipp: Menü „Transformieren > Variable berechnen > Formel MEAN(..., ..., ...)“ und in die Klammer alle Items einfügen [→ umcodierte Items beachten!]
 - Berechnung von **Mittelwerten** (M), **Standardabweichungen** (SD) und **Häufigkeitstabellen** für die erstellten Variablen sowie demografische Größen wie Alter und Geschlecht
 - Eine **Korrelationstabelle** für die zentralen Variablen der Studie erstellen
 - Ermittlung der **internen Konsistenzen** für die verwendeten Fragebögen
 - ✓ Tipp: Menü „Analysieren > Skala > Reliabilitätsanalyse“ und im Fenster alle Items desselben Fragebogens einfügen [→ auf mögliche Umcodierungen achten!]



Exkurs > Die lohnenswerte Arbeit mit Syntax

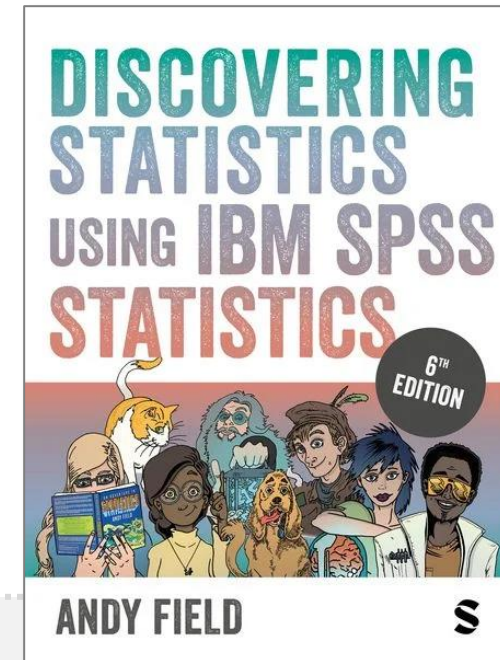
- SPSS bietet mit seiner „**Syntax**“-Funktion ein sehr nützliches Tool:
Alle durchgeführten Schritte können wie ein ‚**Tagebuch**‘ gespeichert werden
 - Ermöglicht später die **Rückverfolgung** bzw. einfache Wiederholung der verschiedenen Klicks in einem Fenster
- Um eine Aktion per Syntax abzuspeichern, wählen Sie am Ende eines SPSS-Menüs **nicht** den Button „OK“, sondern den Button „Einfügen...“
 - Code wird automatisch in das geöffnete Syntaxfile (.sps) eingetragen → bitte abspeichern
- Zum Ausführen der Aktion entsprechenden Code markieren und grünen Pfeil klicken





Next steps > Gut vorbereitet in die Statistik-Sprechstunde

- Welche statistischen Tests zur Prüfung meiner Hypothesen/Forschungsfragen habe ich präregistriert? → eng an dieser Anleitung orientieren
 - Konkrete Fragen und Probleme für Betreuer:in notieren
 - Datensatz ggf. vor dem Gespräch zusenden
- **Immer eine gute Hilfe:** Die **Statistikbibel** von Andy Field



• Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th edition). Sage.





Next steps > Neben der Analyse...

Datenmanagement

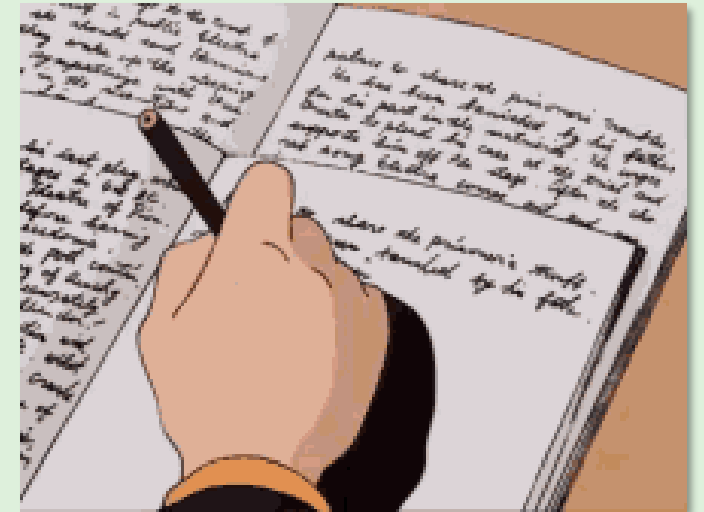
- ✓ Halten Sie **Ordnung** in Ihrem Auswertungsordner...
- ✓ ...denn in vielen empirischen Projekte sollen Daten im **digitalen Anhang** abgegeben werden (z. B. auch BA/MA)
- ✓ Bewahren Sie nur **anonyme** und **keine persönlichen Daten** Ihrer Versuchspersonen auf (Identifizierbare Informationen ggf. aus Datensatz entfernen)

Vergütung organisieren

- ✓ Wenn den Versuchspersonen eine **Vergütung** angekündigt wurde, bitte daran denken!
- ✓ **VP-Stunden** → Herunterladen der Liste an User-Kürzeln (z. B. SoSciSurvey „Getrennt erhobene Kontaktdaten“) und zeitnah per **Mail an Betreuer:in**
- ✓ **Geld/Verlosung/Spende** → proaktiv mit Betreuer:in besprechen

Schreiben, schreiben, schreiben

- ✓ Auch wenn die Datenanalyse noch läuft...





@mp_tuc



@jpstein.de



jpstein@phil.tu-chemnitz.de



Viel Erfolg bei Ihren wissenschaftlichen Vorhaben...

...und entfesseln Sie die inneren Statistikfans!

Prof. Dr. Jan-Philipp Stein