

# Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistics

Charles M. Judd (/search?value1=Charles+M.+Judd&option1=author&noRedirect=true)<sup>1</sup>, Jacob Westfall (/search?value1=Jacob+Westfall&option1=author&noRedirect=true)<sup>2</sup>, and

 View Affiliations

Vol. 68:601-625 (Volume publication date January 2017)

First published as a Review in Advance on September 28, 2016

© Annual Reviews

## ABSTRACT

Traditional methods of analyzing data from psychological experiments are based on the assumption that there is a single random factor (normally participants) and the targets to which they respond, such as words, pictures, or individuals). The application of traditional analytic methods to the data from a comprehensive typology of designs involving two random factors, which may be either crossed or nested, and one fixed factor, condition. We present approaches for power estimation for all designs. We then discuss issues of design choice, highlighting power and feasibility considerations. Our goal is to encourage appropriate design choices for participants and targets.

## Keywords

**experimental design** (/search?option1=pub\_keyword&value1="experimental design"), **mixed models** (/search?option1=pub\_keyword&value1="mixed models"), **random factors** (/search?option1=pub\_keyword&value1="statistical power"), **sample size** (/search?option1=pub\_keyword&value1="sample size"), **effect size** (/search?option1=pub\_keyword&value1="effect size")

## INTRODUCTION

Psychologists learn early in their statistical training to use analysis of variance procedures ( $t$ -tests and ANOVA) to analyze data from designs in which participants are a random factor, whereas participants are a random factor, meaning that the participants used in any particular study are thought to be a sample of participants that differ in their mean condition difference, as well as an estimate of the uncertainty surrounding that difference, by examining the variability across participants (i.e., across conditions). Given the variability of participants, is sufficiently large to permit the belief that it would continue to be found with other samples of participants.

However, many questions in psychology do not lend themselves easily to these well-learned analytic approaches. Frequently, research questions demand effects that should be sought. For instance, a memory researcher might be interested in memory for word lists under different conditions and wish to reach conclusions that might have been used. Likewise, a social psychologist might ask participants to respond to faces of individuals coming from two different ethnic or racial categories. Other samples of faces that might have been used. Additionally, consider a clinical psychologist who is interested in showing that a new therapeutic approach is superior to data from patients who are being treated by therapists under either the new or the standard approach. Again, generalization of any differences should reasonably be expected.

Because psychological researchers are not routinely trained in the analysis of data from designs, such as those just illustrated, that have multiple random factors, ignoring one of the random factors so that the familiar  $t$ -tests and ANOVA procedures can be used. For instance, the memory researcher would typically compare means across conditions, compute, for each participant, means across faces within a racial category; and the depression researcher might simply ignore the therapists in the analysis. This is inappropriate because they have been shown to result in seriously inflated type I statistical errors, leading researchers to claim statistically significant effects (Judd et al. 2012). Many failures to replicate experimental results likely stem from this (Westfall et al. 2015).

To remedy these errors, in this review we provide a thorough treatment of the design and analysis of experiments in psychology that have more than one random factor. If an effect arises from a single random factor (e.g., participants), there exist multiple sources of error variation arising from multiple random factors (e.g., words as well as participants). A more general analytic approach is necessary, in which those multiple sources of random variation are explicitly modeled and estimated. This more general analytic approach (Westfall et al. 2012). We provide a thorough treatment of this approach in the context of psychological experimental designs having two random factors.

We begin with the familiar designs involving only one random factor, participants, and a single fixed condition factor having two levels. These are the experimental designs that these procedures can be recast into the mixed-model framework so that the familiar analyses become special cases of mixed-model analyses.

We then turn to designs having two random factors (which we call participants and targets) and one fixed factor (which we call condition), and present the comprehensive typology of all such designs, including designs in which the two random factors are crossed and designs in which one random factor is nested. We estimate that is modeled on Cohen's  $d$  (i.e., the standardized mean difference; **Cohen 1988**) but generalized to the current designs involving two random factors. Next we develop procedures for the estimation of statistical power in the context of the designs considered, including providing access to a web-based application, choices, and the efficiency of alternative designs.

In the concluding sections of the review, we expand the design possibilities, discussing designs with more than two levels of condition, with multiple fixed factors, and designs with multiple random factors.

## MIXED MODELS FOR DESIGNS WITH ONE RANDOM AND ONE FIXED FACTOR

We begin with familiar designs in which there is one fixed factor, condition, having two levels, and only one random factor, participants. For instance, imagine participants under two conditions, with and without stress. In this context, there are two possible designs: participants are in both conditions or participants are in only one condition. The first design is called a between-participant design (**Smith 2014**). We refer to the first design as the C design, meaning that participants are crossed with condition, and the second design as the N design, meaning that participants are nested within condition. The standard least-squares analysis for data from the C design is the paired  $t$ -test or, equivalently, a repeated-measures ANOVA. For data from the N design, the standard least-squares analysis is the one-way ANOVA. To recast the analysis of data from these designs into the mixed-model terminology, we first specify the possible sources of variation in the observations from the C design. We refer to the individual participant and  $k$  to the condition under which the observation is taken. The mixed-model specification of the individual observation is

$$Y_{ik} = \beta_0 + \beta_1 c_k + \alpha_i^P + \alpha_i^{P \times C} c_k + \varepsilon_{ik}.$$

The values of  $c_k$  represent condition and are assumed to be contrast- or deviation-coded <sup>1</sup> (i.e.,  $c_1 = 1$  and  $c_2 = -1$ ). The terms  $\beta_0$  and  $\beta_1$  represent the fixed effects of condition. In mixed-model terminology,  $\beta_0$  is the fixed intercept and  $\beta_1$  is the fixed slope of condition. What makes this a mixed model is that, in addition to these fixed effects, there are random components of variation that vary across the participants in the design. The following are the random components of variation in the observations:

$$\text{var}(\alpha_i^P) = \sigma_P^2, \quad \text{var}(\alpha_i^{P \times C}) = \sigma_{P \times C}^2, \quad \text{var}(\varepsilon_{ik}) = \sigma_E^2.$$

The variance attributable to participant mean differences is designated as  $\sigma_P^2$ . In the language of mixed models, this is the random variation across participants (i.e., the variance of the random intercepts). The variance of the participant-by-condition interaction effects (i.e., the variance of the random slopes) is  $\sigma_{P \times C}^2$ . In the language of mixed models, this is the random variation across participants in their condition slopes. This is also in the standard ANOVA approach to these designs; the mixed-model specification makes them explicit. Additionally, in the mixed-model specification, we allow for the possibility of a covariance between the random intercepts and slopes, allowing those participants with higher mean responses to have smaller or larger condition differences. This covariance is typically ignored in the standard ANOVA approach.

$$\text{cov}(\alpha_i^P, \alpha_i^{P \times C}) = \sigma_{P, P \times C},$$

allowing those participants with higher mean responses to have smaller or larger condition differences. This covariance is typically ignored in the standard ANOVA approach.

The mixed model given in **Equation 1** can be rewritten to make clear that the  $\alpha_i^P$  and  $\alpha_i^{P \times C}$  terms represent random variation in the intercepts and slopes across participants:

$$Y_{ik} = \underbrace{(\beta_0 + \alpha_i^P)}_{\text{intercepts}} + \underbrace{(\beta_1 + \alpha_i^{P \times C})}_{\text{slopes}} c_k + \varepsilon_{ik}.$$

Cast this way, we have a linear model with a single predictor variable,  $c_k$ , specifying varying intercepts and slopes over and above their fixed (or average) components.

As already specified, the condition effect in the above model is captured by  $\beta_1$ , which equals  $(\mu_1 - \mu_2)/2$ . **Cohen (1988)** defined the general standardized effect size  $d$  as the difference between the mean observations within the conditions: <sup>2</sup>

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_P^2 + \sigma_{P \times C}^2 + \sigma_E^2}}.$$

This full model, with all the random components of variation, is estimable only when each participant is crossed with condition (as in the C design) and when there is more than one replicate per participant per condition. In the C design with only one replicate (i.e., one observation from each participant in each condition) and in the N design, one can still estimate the condition effect, but one cannot estimate the random components of variation. We do not consider in detail designs with multiple replicates (although see the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev.ps.2014.01.01>)). The specifications for the C and N designs become a simple matter of trimming from the full model.

One important issue in estimating the mixed model is the structure of the data file. In the typical ANOVA approach to data, each participant has one row of data for each condition. In the mixed-model approach, each row of data represents an observation taken from a particular participant in a particular condition. For instance, if a given participant were to be observed in both conditions with three replicates each, the data file would have six rows of data for that participant.

The code for estimating the mixed model specified above for these data is as follows:

```

SAS:   proc mixed;
        class participant;
        model y = c;
        random intercept c/sub = participant type = un;
        run;
SPSS:  mixed y with c
        /fixed = c
        /print = solution testcov
        /random = intercept c | subject(participant) covtype(un).
R:     model <- lmer(y ~ c + (c|participant))

```

In each case, the fixed effects are specified in the mixed model, modeling the observations as a function of condition. Implicit in the model specification are random components of variance, indicating that both the intercept and the slope for condition are allowed to vary randomly across participants. In the *lme4* package, the slope for *c* (and implicitly the intercept) varies across participants. The “un” option in both SAS and SPSS specifies that the random intercept and slope are uncorrelated. The output provides slope fixed estimates (along with standard errors) and the variances and covariance of the random intercepts and slopes. Assumptions are that the random effects are independent (no carryover or lagged effects).

We turn now to the C and N designs with a single replicate. In these designs, as we have said, the same underlying components of variance contribute to the observations.

### Mixed-Model Specification of the C Design

In the C design, with each participant in both conditions, the same fixed effects can be estimated. However, not all of the random components are estimable from the observations; the random intercept and slope cannot be disentangled from the residual error term  $\sigma_E^2$ . Accordingly, in the mixed model code, one simply eliminates the random variance components for the intercept and slope. Thus, in this design, one estimates only two random variance components, participant intercepts and residual error.

The general effect size for this design is given as

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_P^2 + [\sigma_E^2 + \sigma_{P \times C}^2]}}$$

The denominator of this effect size contains, as before, all three random sources of variation in the observations, but in this case two of these sources are eliminated, leaving only the variance due to participant slopes.

The test of the condition effect is based on a *t*-statistic that divides the estimated mean difference between the conditions by its estimated standard error. In this design, the standard error includes the participant slope variance and the residual error variance. The variance attributable to participant intercepts (or means) does not contribute to the standard error. The operative effect size is the mean condition difference divided only by those variance components that contribute to its standard error. Accordingly, for the C design, the operative effect size is

$$d_0 = \frac{\mu_1 - \mu_2}{\sqrt{[\sigma_E^2 + \sigma_{P \times C}^2]}}$$

In any sample of data, the operative effect size is estimated as the mean observed condition difference divided by the square root of the estimated residual error variance. The general effect size, rather than operative, effect size is typically reported. We give both to clarify those variance components that do and do not contribute to the standard error of the condition difference.

### Mixed-Model Specification of the N Design

In this design—the classic two-group between-subjects design—each participant is observed in only one condition. As a result, the error variance contains all random components. Accordingly, in the mixed-model specification, no random components are estimable except for residual error. In the computer code to estimate and test the condition difference, one simply eliminates the random variance components for the intercept and slope.

The general effect size for this design is

$$d = \frac{\mu_1 - \mu_2}{\sqrt{[\sigma_E^2 + \sigma_P^2 + \sigma_{P \times C}^2]}}$$

As in the general effect size for the C design, the brackets indicate that the variance due to participant intercepts and participant slopes is now part of the denominator. The operative effect size is the square root of the estimated residual error variance.

Because variances due to both participant intercepts and participant slopes contribute to the estimated residual error in this design, all three components contribute to the standard error. The operative effect size is identical to the general effect size.

The mixed-model specifications for the C and N designs yield tests of the condition difference that are identical to the comparable standard ANOVA approach. The standard ANOVA approach treats the individual participant as the unit of analysis and does not normally make explicit all sources of variation in the data. The mixed-model approach makes explicit all sources of variation in the data and allows multiple simultaneous sources of random variation in the data. For this reason, the mixed-model approach is appropriate for the analysis of data with two random and one fixed factor.

## DESIGNS WITH TWO RANDOM AND ONE FIXED FACTOR

With only one random factor, the design alternatives are limited. With two random factors, the design possibilities grow considerably. The random factors may also be crossed with or nested within the levels of the fixed factor. In this section, we lay out all the design possibilities. We continue to refer to the fixed factor as the condition. We assume the goal is to estimate and test the condition difference so that inferences can be made to other samples of participants and targets that might be used. We start with the most general design, in which all factors are crossed with each other (every participant responds to every target in both conditions) and in which each target in each condition). We refer to this as the most general design because it is only in the context of this design that we can define and estimate all the variance components. In this design can we give the full mixed-model specification and its associated code for estimation. We then provide a general effect size definition as the magnitude of the condition difference. We turn next to more widely used designs that do not include multiple replicates and in which, therefore, not all of the variance components are estimable. These designs, in which one random factor is nested within the other. Accordingly, the designs that we consider, and their mixed-model specifications, bridge two rather distinct design types that are commonly referred to as multilevel or hierarchical linear models (Hox 2010, Raudenbush & Bryk 2002, Snijders & Bosker 2011).

As was the case for the specific designs with participants as the only random factor that were considered in the previous section, these specific designs differ from the most general design in that not all variance components that contribute to the observations can be estimated. For each design, we give those variance components that are estimable and those that are not and then provide the code for estimating condition differences with generalization across both participants and targets. For each design, we also give appropriate design-specific effect sizes.

### Mixed-Model Specification and Effect Size for the Most General Design

In this section, we present the full mixed-model specification for designs with the two random factors of participants and targets and the fixed factor of condition. We start with the most general design in which all three factors are fully crossed and in which there are multiple replicates. This is the most general design in the sense that only in this design can we estimate all the variance components. We then represent modifications of this design in which some of the observations are systematically missing and, accordingly, in which some of the variance components are not estimable. We assume a single dependent variable with variation accruing from a condition difference; a series of random effects attributable to the underlying factors. The response of a participant to the  $j^{\text{th}}$  target in the  $k^{\text{th}}$  condition is

$$Y_{ijk} = \beta_0 + \beta_1 c_{ijk} + \alpha_i^P + \alpha_i^{P \times C} c_{ijk} + \alpha_j^T + \alpha_j^{T \times C} c_{ijk} + \alpha_{ij}^{P \times T} + \alpha_{ij}^{P \times T \times C} c_{ijk} + \epsilon_{ijk}$$

and the following are the sources of variation in  $Y_{ijk}$ :

$$\begin{aligned} \text{var}(\alpha_i^P) &= \sigma_P^2, & \text{var}(\alpha_i^{P \times C}) &= \sigma_{P \times C}^2, & \text{cov}(\alpha_i^P, \alpha_i^{P \times C}) &= \sigma_{P, P \times C}, \\ \text{var}(\alpha_j^T) &= \sigma_T^2, & \text{var}(\alpha_j^{T \times C}) &= \sigma_{T \times C}^2, & \text{cov}(\alpha_j^T, \alpha_j^{T \times C}) &= \sigma_{T, T \times C}, \\ \text{var}(\alpha_{ij}^{P \times T}) &= \sigma_{P \times T}^2, & \text{var}(\alpha_{ij}^{P \times T \times C}) &= \sigma_{P \times T \times C}^2, & \text{cov}(\alpha_{ij}^{P \times T}, \alpha_{ij}^{P \times T \times C}) &= \sigma_{P \times T, P \times T \times C}, \\ & & & & \text{var}(\epsilon_{ijk}) &= \sigma_E^2. \end{aligned}$$

As above,  $\beta_0$  and  $\beta_1$  in this model represent the fixed effects and capture, respectively, the overall mean response and the condition difference in responses. The random effects are given intuitive interpretations in **Table 1**. To show more clearly the specification of some of these components as random intercept components and others as random slope components, we can rewrite the model as

$$Y_{ijk} = \underbrace{(\beta_0 + \alpha_i^P + \alpha_j^T + \alpha_{ij}^{P \times T})}_{\text{intercepts}} + \underbrace{(\beta_1 + \alpha_i^{P \times C} + \alpha_j^{T \times C} + \alpha_{ij}^{P \times T \times C})}_{\text{slopes}} c_{ijk} + \epsilon_{ijk}$$

On the basis of this model and again using **Cohen's (1988)** specification of the effect size, the following can be defined as the general effect size for this design:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_P^2 + \sigma_{P \times C}^2 + \sigma_T^2 + \sigma_{T \times C}^2 + \sigma_{P \times T}^2 + \sigma_{P \times T \times C}^2 + \sigma_E^2}}$$

For mixed-model estimation, the data file is again structured so that each individual observation is a row of data. The code for estimating effects for data from this design is provided in the **Code** section.

**Table 1**  
Definitions of random variance and covariance components in the designs considered in this review

Toggle display: Table 1 

[Open Table 1 fullscreen](#)

Variance or covariance component		
$\sigma_P^2$		Participant
$\sigma_{P \times C}^2$		Participant
$\sigma_{P, P \times C}$		Participant condition d
$\sigma_T^2$		Target inter

Variance or covariance component		
	$\sigma_{T \times C}^2$	Target slopes
	$\sigma_{T, T \times C}$	Target intercept differences
	$\sigma_{P \times T}^2$	Participant-toward-participant differences
	$\sigma_{P \times T \times C}^2$	Participant-by-target interaction
	$\sigma_{P \times T, P \times T \times C}$	Participant-by-target interaction (also show I)
	$\sigma_E^2$	Residual error

Variables: C, fixed condition factor; E, error; P, random participant factor; T, random target factor.

```

SAS:  proc mixed;
      class participant target;
      model Y = c;
      random intercept c/sub = participant type = un;
      random intercept c/sub = target type = un;
      random intercept c/sub = participant*target type = un;
      run;

SPSS:  mixed y with c
      /fixed = c
      /print = solution testcov
      /random = intercept c | subject(participant) covtype(un)
      /random = intercept c | subject(target) covtype(un)
      /random = intercept c | subject(participant*target) covtype(un).

R:  model <- lmer(y ~ c + (c|participant) + (c|target)
      + (c|participant:target))

```

This code is an extension of the code given above for designs with one random factor (see Mixed Models for Designs with One Random and One Fixed Factor) and includes intercepts and slopes due to participants, those due to targets, and finally those due to the participant-by-target interaction.

As before, not all variance components contribute to the standard error used to test the condition difference in this design. Accordingly, the operative effect size is the standard error of those components that contribute to its standard error, is:

$$d_0 = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_{P \times C}^2 + \sigma_{T \times C}^2 + \sigma_{P \times T \times C}^2 + \sigma_E^2}}$$

In the following sections, we systematically define the possible designs that involve two random factors (participants and targets) and a single fixed factor (condition). These designs are a subset of the most general design considered above but with systematically missing observations. Each design provides an estimate of the fixed effects of interest, but some of these components are confounded with each other, and thus model specification and effect sizes must be tailored to each particular design.


## Design Possibilities

To define the full range of designs that have the three factors of condition, participants, and targets, we must consider the three possible pairs of these factors. For each pair, two factors may be crossed or nested. We use C and N to indicate whether the factors in each pair are crossed or nested, respectively. Each design is thus identified by three letters: the first C or N indicates whether targets are nested within condition or crossed with condition; the second C or N indicates whether targets are crossed with condition or nested within condition; and, finally, the third letter defines whether participants are nested within condition or crossed with condition. If the two random factors are nested, there are two possibilities: Either targets are nested within participants (meaning that each participant responds to a unique set of targets) or participants are nested within targets (meaning that each target is responded to by a unique set of participants). In the first case, the final letter in the definition of each design is N<sub>P</sub>, meaning that participants are the higher-level factor with respect to targets; in the second case, the final letter in the definition of each design is N<sub>T</sub>.

The designs are listed in **Table 2**; each design is identified by the labels defined above. We now further define and illustrate each of these designs. We start with each other.

**Table 2**  
Typology of designs with two random factors [participants (P) and targets (T)] and one fixed factor (condition)

Toggle display: Table 2  

[Open Table 2 fullscreen](#) 

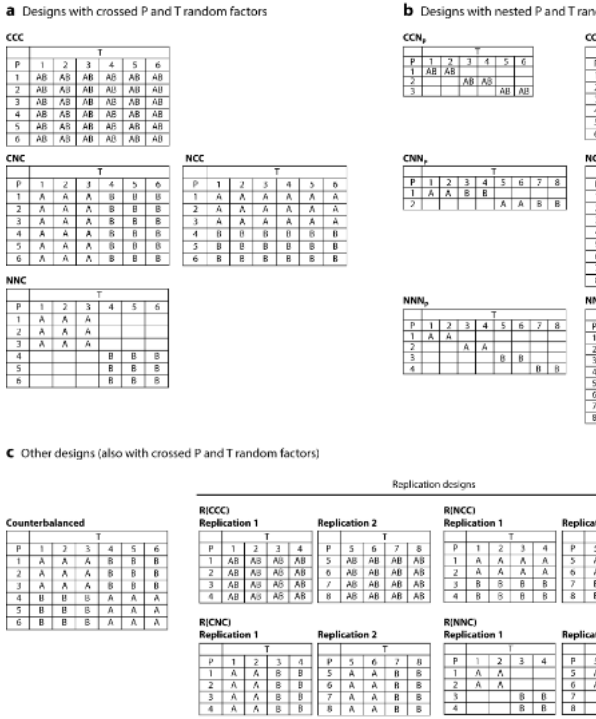
How are P and T related to condition?	How are P and T related?	
	P and T crossed	T nested in P
P and T crossed with condition	CCC	CCN <sub>P</sub>
P crossed with condition; T nested in condition	CNC	CNN <sub>P</sub>
P nested in condition; T crossed with condition	NCC	Impossible
P and T nested in condition	NNC	NNN <sub>P</sub>

**Designs with crossed random factors.**

The four cells in the first column of **Table 2** define four designs with the final letter C. These designs are illustrated in **Figure 1 a**. For ease of depiction, in the likely be in any actual study.

**Figure 1**

Illustrative matrices for all designs having two random factors, participants (P, rows) and targets (T, columns), and one fixed factor, condition, with two levels (A and B) under which particular observations



Judd CM, et al. 2017. Annu. Rev. Psychol. 68:601–25

[Click to view](#)

[\(/docserver/fulltext/psych/68/1/ps680601.f1.gif\)](#) Download as PowerPoint [\(/docserver/fulltext/psych/68/1/ps680601.f1.pdf\)](#)

The first design, CCC, is the fully crossed design in which every participant responds to every target twice, once in each condition. Imagine a design in which faces, one version morphed towards a prototypic White face and the other morphed towards a prototypic Black face. Faces constitute the random target factor and condition is the fixed factor. Each face is judged in both its White-morphed version and its Black-morphed version (i.e., condition).

The CNC design is one in which every target is responded to by every participant, but each target is in only one condition. Imagine a variation of the previous design in which actual faces of White and Black individuals rather than morphed versions of faces. Thus, each individual face is either White or Black, so targets are nested within condition. Each target is judged by half White and half Black.

In the NCC design, participants are nested within condition and targets are crossed with condition. Imagine that participants complete a series of target judgments under both load conditions, albeit by different participants.

In the NNC design, both random factors are nested within condition. Imagine that participants make career likelihood judgments of faces (e.g., “How likely is this person to be a doctor?”). Each target is judged only by one set of participants. Gender of target is the condition variable of interest.

**Designs with nested random factors.**

Designs in the second and third columns of **Table 2** have one of the two random factors nested within the other. In the second column, in which targets are nested within condition, every target is responded to by its own unique set of participants. These designs are illustrated by the matrices in **Figure 1b**.

The CCN<sub>P</sub> and CCN<sub>T</sub> designs have one random factor nested within the other, but both of these factors are crossed with condition. The classic nested design is the CCN<sub>P</sub> design. In one version of this design, the instructors evaluate their students in two different conditions or subjects, math and language. The question is whether the evaluations depend on the subject matter. In the CCN<sub>T</sub> design, the students are now the participants and they evaluate their instructors, but these two groups have switched their roles in terms of the design: The instructors elicit responses, and we thus designate them as the target. Both random factors are crossed with condition.

In the case of nested random factors where one of the factors is crossed with condition and the other is nested within condition, the higher-order random factor of the cells in the second and third columns of **Table 2** define impossible designs. In the  $CCN_p$  design, targets are nested within participants, participants are each asked to nominate and judge their two closest male and two closest female friends. Thus, each participant has a unique set of targets (friends) who are given different ratings to their nominated male friends than their female friends.

In the  $NCN_T$  design, participants are nested within targets, targets are crossed with condition, and participants are nested within condition. In this case, the common (male) friend, the target. Participants (those who do the rating of their common nominating friend) are now nested within gender (their own), but targets are still nested within condition.

The final two designs of **Table 2** are the fully nested designs,  $NNN_p$  and  $NNN_T$ , in which either targets are nested within participants or the other way around. In the  $NNN_p$  design, male and female participants are recruited, and they nominate and rate as targets two friends of only their own gender. In this case, targets are nested within participants, and participants are nested within condition. In the  $NNN_T$  design, again imagine that people nominate their two friends, who are, again, the same gender as the nominating person. However, this time, the targets are nested within condition. Participants are now nested within targets and both are nested within condition (gender).

In addition to these designs, there are two final designs, used with some frequency, in which participants and targets are in fact confounded, with just a single target. Imagine research in which each participant thinks of a single friend and rates him or her, either in one condition only or in both. Thus participant and target are confounded. The analysis of this design is formally equivalent to those with one random factor that we considered above. However, in this case the random factor is not participant, but participant and target, as well as their interaction.

### Other designs.

**Table 2** provides a coherent way of defining the possibilities with two random and one fixed, two-level factor. However, other possibilities deserve discussion.

First, there is a variation on the fully crossed CCC design that we call the counterbalanced design (**Westfall et al. 2014**). This is a fully crossed design in the sense that all conditions, participants, and targets are crossed. Unlike the CCC design, however, each participant responds to each target in just one condition. As shown in **Figure 1c**, participants and targets are fully crossed, but condition is confounded with target. In the CCC design, condition, participants, and targets are fully crossed, whereas in the counterbalanced design, condition is confounded with target. In some problems, some while under cognitive load and others without load. Every participant does all problems, but the division of the problems between the half of the problems is counterbalanced.

Second, there are four designs that we refer to as replication designs in that they replicate some of the designs of **Table 2** with multiple sets of participants and targets (i.e., multiple replications of the same participant, target, and condition). We mean something entirely different by replication designs, i.e., that an entire previously defined design is replicated multiple times. The first row of **Table 2**, in which both participants and targets are crossed with condition. Suppose that, rather than fully crossing participants and targets, we group participants and targets into multiple groups. In each group, participants and targets are fully crossed, but there are multiple such groups. This design essentially replicates the CCC design many times, with each replication having its own set of participants and targets. Again, a replication is defined as a specific group or subset of participants and targets. In **Figure 1c** we have illustrated the R(CCC) design with the number of replications. For example, suppose participants are put in groups of four and everyone in a particular group responds to the same four targets twice, once in one condition and once in the other. This design has eight replications. The advantage of this design over the fully crossed design is that it potentially reduces participant load (i.e., participants do not need to respond to all targets in all conditions) for statistical efficiency reasons considered below (see Power Considerations and Research Design).

The R(CCC) design is the replication design from the first row of **Table 2**. The other three replication designs correspond to the remaining three rows of **Table 2**. The R(NCC) design is the NCC design replicated multiple times with different sets of participants and targets; each target occurs in only one condition or the other. The R(NCC) design is the NCC design with multiple replications of different sets of participants and targets. And finally, the R(NNC) design is the NNC design with multiple replications of different sets of participants and targets.

These replication designs, with participants crossed with targets in each replication, become the nested designs of the second and third columns of **Table 2** when the number of replications is one. The R(CCC) design becomes the  $CCN_p$  design if each replication contains only a single participant, responding to the targets that are unique to that replication; if there are multiple participants in that replication. The other replication designs also become the nested designs of the third and fourth columns of **Table 2** when the number of replications is one.


### Design-Specific Estimation and Effect Sizes


In this section, we discuss the mixed-model specification that estimates the condition difference given all of the random variance components that are estimated. These variance components, that, along with residual error, contribute to the total variation in observations. These are defined in **Table 1**. In the fully crossed design with multiple replications, all of these variance components are estimable. Accordingly, we gave the mixed-model code in SAS, SPSS, and R; this code specifies how one estimates the condition difference and the effect sizes for this design; the general effect size is defined as the mean condition differences divided by all six variance components plus the residual variance, and the specific effect sizes are defined as the mean condition differences divided by the variance components that contribute to the standard error of the condition difference.

In the second column of **Table 3**, we present the general effect sizes for all of the designs that we have defined. (The third column of this table lists the non-estimable effect sizes for the designs that are discussed in the section below devoted to that subject.) Consistent with our earlier treatment of designs that have participants as the only random factor, the confounding of condition and target in the  $CCN_p$  design, the denominators of the general effect sizes include, for all designs, all six variance components defined in **Table 1** plus random error variance, but many of the

indicates the component that is estimable in the mixed-model specification, and the components that follow within the brackets are those that are confounded with the component that is estimable. For example,  $\sigma_{P \times T \times C}^2$  indicates that the variance component for the participant-by-target-by-condition interaction is confounded with the residual error variance. The **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>). All the information necessary for specifying the mixed-model specification for each design is provided in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>), we give the code (again in SAS, SPSS, and R) for each design. The general effect size in **Table 3**. The rule is that one specifies as random effects those variance components that are not contained in brackets in the denominator of the effect size. Variance components that are in brackets are not included in the model by default and does not need to be specified explicitly.)

**Table 3**  
General effect sizes ( $d$ ) and noncentrality parameters for the designs of **Table 2** and **Figure 1**<sup>a</sup>

Toggle display: Table 3  ▼

[Open Table 3 fullscreen](#) 

Designs	
CCC	
CNC	
NCC	
NNC	
CCN <sub>p</sub>	
CCN <sub>T</sub>	
CNN <sub>p</sub>	
NCN <sub>T</sub>	
NNN <sub>p</sub>	
NNN <sub>T</sub>	
Counterbalanced	
R(CCC)	Same as CCC
R(CNC)	Same as CNC
R(NCC)	Same as NCC
R(NNC)	Same as NNC

<sup>a</sup>Brackets indicate the confounding of variance components. All variance components in the noncentrality parameters are defined in **Table 1**. The number of participants is  $p$  and the number of targets is  $q$ . In replication designs, the number of replications is  $r$ .

In the following paragraphs we provide illustrations for a few of the designs of how one goes from the general effect sizes in **Table 3** to the mixed model code. For each design, we provide the mixed model code (in SAS, SPSS, and R) for each design. We also briefly discuss for each design the estimable components that do not contribute to the standard error of the condition difference. The general effect sizes (which are listed in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>)).

The first design we consider is the CCC design. The only use of brackets in the denominator of its general effect size in **Table 3** is  $[\sigma_E^2 + \sigma_{P \times T \times C}^2]$ , which indicates that the variance component for the participant-by-target-by-condition interaction is confounded with the residual error variance. Accordingly, in modifying the code given in the section Mixed-Model Specification and Effect Size for the Most General Design (the most general design is the CCC design), we must specify random condition slopes for the participant-by-target interaction. Because participants and targets (and their interaction) are crossed with condition, the implicit residual error term is confounded with the condition difference, although those components are estimable and should be included in the model.

Second, the NCC design has two sets of brackets in the denominator of its effect size. Variance attributable to the participant by condition interaction is confounded with the residual error variance. Thus, the code must be modified to estimate only random participant intercepts and random condition difference in this design except that due to target intercepts.

As a third example, the NCN<sub>T</sub> design has three estimable components, those due to target variance, target by condition variance, and residual variance. Thus, the implicit residual error term. Target intercept variance, although estimable, does not contribute to the standard error of the condition difference.



At the bottom of the first column of **Table 3**, we indicate that the effect sizes for the four replication designs are identical to those of the parallel designs in all targets). Thus, for instance, the effect size for the R(CCC) design is identical to that given for the CCC design. The syntax for these designs is also the same as for the parallel designs, with the condition-by-replications interaction treated as a fixed factor (along with the fixed condition-by-replications interaction).

We end this section with a final warning on model specification. We have seen published analyses of designs with crossed random factors of participants and targets within participants (e.g., **Toma et al. 2012**). Typically in diary studies, for instance, days of measurement are crossed with participants but are treated as nested random factors. This has been used in the literature for some time, whereas models for crossed random factors are a more recent development. The misspecification of a crossed design risks serious inflation of type I errors if in fact there is nonzero target variance (**Judd et al. 2012**). The lesson is that model specification should follow from theory.

## STATISTICAL POWER FOR DESIGNS WITH TWO RANDOM FACTORS

In this section and in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>), we provide the details of the designs that we have covered. We discuss the general approach to power estimation and then provide a web-based application that computes power for all designs that involve two crossed random effects. The current application (located online at [http://jakewestfall.org/two\\_factor\\_power/](http://jakewestfall.org/two_factor_power/) ([http://jakewestfall.org/two\\_factor\\_power/](http://jakewestfall.org/two_factor_power/))) computes power for both crossed and nested random effects, as well as the replication designs.

Our approach to statistical power estimation is consistent with the general approach laid out by **Cohen (1988)**. One begins by specifying both a null hypothesis and an alternative hypothesis of some magnitude. Power is defined as the probability of correctly rejecting the null hypothesis when the alternative hypothesis is correct. To compute power, we need to know the expected difference for each design; these are given in the denominators of the design-specific operative effect sizes (see the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>)). These are then weighted appropriately by the sample sizes involved in the prospective study [total numbers of participants ( $p$ ) and targets ( $q$ ) in the design and noncentrality parameter] and the hypothesized true effect, which is presented in the third column of **Table 3** for each design. One can think of the noncentrality parameter as approximately equal to the square of the numerator of the noncentrality parameter can be thought of as the expected standard error of the condition mean difference. When squared and multiplied by the number of observations, it is as the expected mean square (EMS) for the condition factor (**Winer 1971**).

Given degrees of freedom for this noncentrality parameter, power can be computed by examining areas under the noncentral  $t$ -distribution, given the multiple comparisons adjustment. If the parameters pool or combine various relevant variance components, the degrees of freedom of the noncentral  $t$  must be approximated. We use the Satterthwaite approximation. Expressions for the approximate degrees of freedom for each design are given in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev-psych-122414-033702>)).

We provide a web-based application ([http://jakewestfall.org/two\\_factor\\_power/](http://jakewestfall.org/two_factor_power/) ([http://jakewestfall.org/two\\_factor\\_power/](http://jakewestfall.org/two_factor_power/))) that computes power for a given design, number of participants, number of targets, <sup>6</sup> the hypothesized mean difference or effect size, and the relevant variance components. In the application, the user has a choice between two different input methods. The first method requires the user to input the mean difference expected and estimated values for all of the estimable variance components. An often-simpler option is to input what might be thought of as the relative magnitude of the estimable variance components for each design (the proportion of the total variance in the observations attributable to a particular variance component). We use Partitioning Coefficients (**Goldstein et al. 2002**), and designate them as  $V$  (e.g.,  $V_P$  for participant intercept variance,  $V_{T \times C}$  for target slope variance). By definition, the sum of the  $V$ 's is 1.

## POWER CONSIDERATIONS AND RESEARCH DESIGN

All designs permit an estimate of the condition difference. Therefore, in making a decision about which design to use, the most important considerations are Power and Feasibility Considerations. In this section, we consider those factors influencing the power to detect the anticipated condition difference.

In general, the smaller the variance components that contribute to the noncentrality parameter (or operational effect size) and the larger the relevant sample size, the greater the power. Power is determined by the participant variance components and the participant sample size. In the designs that we are now considering, power is determined by these matter varies from design to design. The important point, however, is that we must think in terms of multiple relevant variance components and multiple comparisons.

To increase power in designs with participants as the only random factor, researchers can either decrease the error variability in the data or increase the number of participants. Both are obvious. Those associated with decreasing participant variability are less obvious. Selecting participants who are relatively homogeneous on the target variable, so, however, restricts one's ability to generalize observed results to other samples that are not so restricted.

The same considerations hold in thinking about designs with multiple random factors, in which variance components due to targets and their sample size, in particular, dramatically increases as the number of targets in a design increases. Additionally, if we restrict the variance attributable to targets through pretesting, removing irrelevant features from target faces in research on face perception to edit target faces to eliminate facial hair and other idiosyncrasies. However, restricting target variance imposes a cost in that one must restrict the number of targets. The same power considerations apply to the sampling of targets as to the sampling of participants. Larger and more homogeneous samples of both increase power.

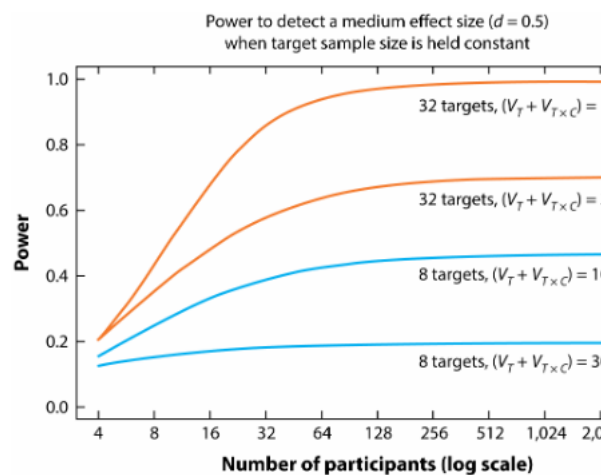
## Increasing the Sample Sizes

In designs with just one random factor, as the participant sample size increases, power eventually approaches one. However, in many of the designs considered, as the sample size of the other random factor increases power, but generally to a limit of less than one. This is a surprising result; in our experience, many researchers do not have adequate power. When the number of targets in a design is small, power will increase as the participant sample size increases but may asymptote at levels that are low. The number of targets used is typically substantially smaller than the number of participants [e.g., a meta-analysis by **Bond & DePaulo (2008)** in one research design]. We think that power is determined only (or primarily) by the sample size of participants.

To illustrate, in **Figure 2** we plot the power to detect a medium effect size of  $d = 0.5$  in the CNC design as a function of number of participants under different conditions. When much variance is due to targets and their interaction with condition, 10% or 30% of the total variance.

**Figure 2**

Plot of statistical power as a function of the total number of participants for the CNC (P and T crossed, P crossed with C, and T nested in C) design. The number of targets has been set to either 8 or 32. The variance components are  $V_P \times C = 0.1$ . Note that these other variance components affect only the rate at which the power functions converge to their asymptotes; they do not affect the maximum attainable power values, which are determined by the number of targets, P, random participant factor; T, random target factor.



Judd CM, et al. 2017. *Annu. Rev. Psychol.* 68:601–25

[Click to view](#)

[\(/docserver/fulltext/psych/68/1/ps680601.f2.gif\)](#) [Download as PowerPoint \(/docserver/fulltext/psych/68/1/ps680601.f2.pdf\)](#)

A similar phenomenon occurs in nested designs (targets within participants or participants within targets), but only for the lower-level factor. For example, in a nested design, the maximum attainable power is less than one if the participant sample size is held constant and the target sample size is increased, but power does approach one as the sample size of the higher-level factor necessarily entails increasing the sample size of the lower-level factor, but the reverse is not true.

Assuming that one is able to vary either the participant sample size or the target sample size (or both), which is expected to have a greater effect on statistical power is the participant and target variance components, and the design of the experiment. The definitive answer is contained implicitly in the noncentrality parameter.

First, assuming crossed random factors, the larger the variance components associated with one random factor (relative to the other random factor), the more power. If the variances are larger than those due to targets, then increasing the sample size of participants reaps greater benefits than increasing that of targets.

Second, if the sample sizes of targets and participants are substantially different, then there will generally be a greater power benefit to increasing the size of the larger sample size than increasing the size of the smaller sample size. For instance, if one sample size is 300 units and the other is 10, then adding an additional 10 units to the larger sample size (for a new total of 310) is more beneficial than adding 10 units to the smaller sample size (for a new total of 20).

Third, all else being equal, it is better to increase the sample size of a random factor that is nested within condition than one that is crossed with condition. The power benefit of a condition difference depends on both the intercept and slope variance components of that factor, whereas when the random factor is crossed with condition, the power benefit is determined by the slope variance component.

Fourth, in a design in which one random factor is nested within the other (e.g., targets within participants), it is usually more effective to increase the sample size of the higher-level factor. As discussed above, the maximum attainable power level when increasing the lower-level sample size in a nested design is, in general, less than 1.0. Accordingly, power increases more quickly by increasing the higher-level sample size than by increasing the lower-level sample size.

### Design Choices: Power and Feasibility Considerations

Power is not the only consideration guiding the choice of design; feasibility issues also figure prominently. We discuss some of those issues in this section.

Although power will often increase dramatically as the target sample size increases, sometimes it is not feasible for participants to respond to a large number of random factors. Researchers often assume that a crossed design is more powerful, but in fact it can be shown that for any crossed design, a nested version of relationships between the random factors and condition, but in which every participant receives different targets—is always more powerful. This difference in number of targets even as the number of responses per participant remains constant.

However, nested designs may in some contexts require unreasonable numbers of targets (the number of participants times the number of responses per participant). For example, then the CCC design involves responses to only 15 targets, whereas the CCN<sub>P</sub> design involves responses to 450 targets, resulting in a potentially dramatic increase in the number of targets to find so many targets. A reasonable alternative is to consider the R(CCC) design, containing, for instance, three replications of the CCC design, with 10 participants per replication; the total number of targets has gone up threefold over the number in the CCC design. Generally speaking, in cases in which each response imposes a considerable burden on participants across different replications of a design, rather than to limit the number of targets by the use of a design in which all participants respond to the same set of targets.

If one has a choice between crossing a random factor with condition and nesting that random factor within condition, then one should always choose to cross the random factor because the variance due to a random factor contributes to the noncentrality parameter (making it smaller) when that factor is nested within condition, but not when it is crossed with condition. It is generally better to cross whichever factor has the larger anticipated variance components. Of course, there are feasibility issues that arise in considering whether to cross or nest, such as the order of and carryover effects, as well as the potential suspicion that participants may develop about the study's purpose. These issues do not arise if the random factor is nested within condition. Finally, if one is using a design with nested random factors and one has the choice of which is the higher-order and which the lower-order factor, then it is always better to nest the lower-order factor within the higher-order factor if the lower-order participants have larger associated variance components than targets, then a design that nests participants within targets is preferable to one that nests targets within participants within the higher-level factor.

## COMPLICATIONS AND EXTENSIONS

Our designs have assumed only two random factors and one fixed factor having only two levels. We have also assumed that when one factor is nested within another, the nesting is random. We first discuss the issue of nonrandom nesting and then turn to design extensions.

### Nonrandom Nesting

When a random factor is nested within condition, differences attributable to that random factor are confounded with condition differences. With random assignment, this is not possible with nonrandom assignment. This is as true of targets as it is of participants. Hence, nonrandom assignment of either participants or targets to conditions is problematic.

When one of the two random factors is nested within the other, either targets within participants or participants within targets, we have assumed random assignment. What if the assignment is nonrandom? For simplicity, we rely on the situation in which targets are nested within participants, but the following considerations apply under the reverse nesting. Nonrandom assignment of targets to participants, resulting in covariances between participant and target intercepts and (perhaps) condition slopes. In many situations, it is likely that targets are assigned to participants with higher means (or that participants have associated higher means). This positive covariance augments the variance components of participants, which generally results in less efficient tests of condition differences. Nonrandom assignment of targets to participants is likely. However, nonrandom assignment of targets to participants does not result in bias in the estimate of the condition difference, so long as participants are randomly assigned to condition. In other words, in a fully nested design, nonrandom assignment of the lower-order random factor to levels of the higher-order one (random assignment of targets to participants) does not bias the condition difference estimate so long as the higher-order random factor is still randomly assigned to condition levels.

In the replication designs, participants and targets are nested within each replication. We have assumed random assignment of both factors to each replication. If the random assignment is nonrandom, then the replications often should be treated as an additional random factor in the design specification rather than as a fixed factor. In replication designs, in consequence, one must have sufficient numbers of replications, because the sample size of this factor now becomes relevant in determining the power of the test.

### Design Extensions

As discussed in the previous section, replications in the replication designs should be considered a random factor in the case of nonrandom nesting within replications. If one presents participants with primes and asks them to respond to subsequently presented targets. In most cases, one should treat both primes and targets, in addition to condition, as random factors. In a review, we can extend the designs and models to cover these scenarios. Additional random factors lead to additional complexities in specifying the random effects model, which grows exponentially if the random factors are all crossed with each other (as in the priming example). Condition slope variance components must be specified in the model, and although the complete model specification is possible, it may be necessary to specify a large number of variance components, leading to possible convergence problems. It is recommended specifying the complete underlying model (including all random variances and covariances that are estimable; see **Barr et al. 2013**), a recommendation that is often difficult to follow. In estimating the model, respecification may help by dropping some of the variance components that represent higher-order interactions that might reasonably be expected to be negligible.

Our designs also have only one fixed factor, condition, with only two levels. In many experiments there are more fixed factors, generally crossed and often with more than two levels. If the effects of those contrasts tested as single-degree-of-freedom tests, additional fixed factors present no further problems other than, again, the complexity of specifying the model. If the random factors that are crossed with those additional fixed factors (and with the interactions of fixed factors).

There may also be continuous covariates that one would like to include in the fixed part of the model. We strongly recommend centering such covariates (Judd & Yzerbyt, 2005) and fixed parameter estimates (including the condition difference).

Finally, we have assumed completely balanced research designs with no missing data. Mixed-model estimation can generally be accomplished with missing data, but this is often done by assuming that missing data are lost, a highly dubious assumption. More detail on missing data is contained in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev.psyc.07.03.08>).

## FROM OUR DESIGNS TO DYADIC DESIGNS

There is extensive literature on what are called dyadic designs (Kenny et al. 2006), in which participants interact with other participants. Dyadic designs are the most natural when the targets are people; nonetheless, even when the targets are inanimate, each observation involves a dyad or a pair.

One advantage of viewing studies with participants and targets as dyadic designs is that there is an established tradition of quantifying the random sources of variance. The relative amount of variance of the different components can be very helpful in planning a study. Designs using the social relations model (SRM) are understood in this way.

The SRM examines observations taken from actors about partners (Kenny et al. 2006). In the parlance of this review, an actor is a participant and a partner is a target. The relationship, i.e., actor×partner interaction. In most applications that use the SRM, there is no fixed variable such as condition, so variances due to condition, actor, partner, and relationship. For instance, Hönekopp (2006) had participants in three studies judge the physical attractiveness of targets' faces using photographs. The variance was due to the participant or actor, 26% due to the target or partner, 33% due to the relationship or participant×target interaction, and the remaining 26% due to error. Knowledge of these variances would prove useful in planning studies on interpersonal attraction.

Although all the designs in this review can be viewed as dyadic designs, only one of the designs is a SRM design: the CCC design, which is called the half-block design because each actor interacts with the same set of partners (i.e., targets). The prototypical SRM design, however, involves a case in which the actors and partners are the same group of people. This design is not considered in this review (see, however, the extensive discussion in Kenny et al. 2006).

Many dyadic designs, like the round robin design, are reciprocal in the sense that observations accrue from both participant A responding to target B and from target B responding to participant A. The CCC design would be a reciprocal design in the following situation: Consider a study in which a sample of men interact with a sample of women; if we were interested in the effect of gender, then we could view the study as a fully crossed design in which, for the male judge condition, men are the participants and women are the targets, and for the female judge condition, women are the participants and the same men are the targets. This SRM design is referred to as the asymmetric block design.

Other designs that we have considered can be made into reciprocal designs by combining them. For instance, in the example we used for the CCN<sub>p</sub> design, the students evaluate their instructor in math and language. If we obtained data from both students and teachers in the same study, we would have a reciprocal one-with-many study. Other designs can also be combined to form reciprocal designs, as is discussed in the **Supplemental Appendix** (<http://www.annualreviews.org/doi/suppl/10.1146/annurev.psyc.07.03.08>).

## CONCLUSION

In psychology experiments, we frequently ask participants to respond to targets (e.g., faces, words, other people) in various experimental conditions. In such cases, it is essential to use other samples of participants and, typically, other samples of targets that might have been used. To permit this, it is essential that the analysis of the results takes into account the variation in targets or ignoring the variation they induce, leads to serious bias and, we suggest, failures to replicate experimental effects when other samples of targets are used.

In this review, we provide an exhaustive typology of designs based on the nesting or crossing of three factors (participants, targets, and condition) in such experiments. We provide estimates of condition effects while treating both participants and targets as random factors. Additionally, we provide tools to estimate the effect size and statistical power for these designs.

We conclude by emphasizing the importance of considering targets as well as participants in determining statistical power. We also discuss considerations to choose appropriate designs and analytic models that incorporate target variation and thus permit conclusions that are more likely to replicate with other samples of targets.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

### ACKNOWLEDGMENTS

We thank Markus Brauer, Leonard Katz, Reinhold Kliegl, Gary McClelland, Dominique Muller, and Vincent Yzerbyt for their helpful comments on drafts of this review.

## LITERATURE CITED

Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed-random effects for subjects and items. *J. Mem. Lang.* 59:390–412 [Google Scholar] (<http://scholar.google.com/scholar?btnG=Search&btnI=Images&btnS=Books&btnL=Library&btnC=Context&btnA=All&btnP=PDF&btnD=Data&btnV=Video&btnO=Other&btnM=More&btnR=Random&btnF=For&btnS+and+items&author=RH+Baayen&author=DJ+Davidson&author=DM+Bates&journal=J.+Mem.+Lang.&volume=59&pages=390-412&pubdate=2008>)

Barr DJ, Levy R, Scheepers C, Tily HJ. **2013**. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68:255–78 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Random+effects+structure+for+confirmatory+hypothesis+testing%3A+keep+it+maximal&author=DJ+Barr&author=R+Levy&author=C+Scheepers&author=HJ+Barr&publication\\_year=2013](http://scholar.google.com/scholar_lookup?title=Random+effects+structure+for+confirmatory+hypothesis+testing%3A+keep+it+maximal&author=DJ+Barr&author=R+Levy&author=C+Scheepers&author=HJ+Barr&publication_year=2013))

Bates D, Kliegl R, Vasishth S, Baayen H. **2015**. Parsimonious mixed models. arxiv:1506.04967v1 [stat.ME]

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR. et al. **2009**. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24:3127–31 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Generalized+linear+mixed+models%3A+a+practical+guide+for+ecology+and+evolution&author=BM+Bolker&author=ME+Brooks&author=CJ+Clark&author=SW+Geange&author=JR+Poulsen&author=BM+Bolker&publication\\_year=2009](http://scholar.google.com/scholar_lookup?title=Generalized+linear+mixed+models%3A+a+practical+guide+for+ecology+and+evolution&author=BM+Bolker&author=ME+Brooks&author=CJ+Clark&author=SW+Geange&author=JR+Poulsen&author=BM+Bolker&publication_year=2009))

Bond CF, DePaulo B. **2008**. Individual differences in judging deception: accuracy and bias. *Psychol. Bull.* 134:477–92 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Individual+differences+in+judging+deception%3A+accuracy+and+bias&author=CF+Bond&author=B+DePaulo&journal=Psychol.+Bull.&volume=134&pages=477-92](http://scholar.google.com/scholar_lookup?title=Individual+differences+in+judging+deception%3A+accuracy+and+bias&author=CF+Bond&author=B+DePaulo&journal=Psychol.+Bull.&volume=134&pages=477-92))

Clark HH. **1973**. The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verb. Learn. Verb. Behav.* 12:335–59 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=The+language-as-fixed-effect+fallacy%3A+a+critique+of+language+statistics+in+psychological+research&author=HH+Clark&journal=J.+Verb.+Learn.+Verb.+Behav.&volume=12&pages=335-59](http://scholar.google.com/scholar_lookup?title=The+language-as-fixed-effect+fallacy%3A+a+critique+of+language+statistics+in+psychological+research&author=HH+Clark&journal=J.+Verb.+Learn.+Verb.+Behav.&volume=12&pages=335-59))

Cohen J. **1988**. *Statistical Power Analysis for the Behavioral Sciences* Hillsdale, NJ: Erlbaum [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Statistical+Power+Analysis+for+the+Behavioral+Sciences&author=Cohen+J&publisher=Erlbaum](http://scholar.google.com/scholar_lookup?title=Statistical+Power+Analysis+for+the+Behavioral+Sciences&author=Cohen+J&publisher=Erlbaum))

Goldstein H, Browne W, Rasbash J. **2002**. Partitioning variation in multilevel models. *Underst. Stat.* 1:223–31 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Partitioning+variation+in+multilevel+models&author=H+Goldstein&author=W+Browne&author=J+Rasbash&journal=Underst.+Stat.&volume=1&pages=223-31](http://scholar.google.com/scholar_lookup?title=Partitioning+variation+in+multilevel+models&author=H+Goldstein&author=W+Browne&author=J+Rasbash&journal=Underst.+Stat.&volume=1&pages=223-31))

Hönekopp J. **2006**. Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *J. Exp. Psychol. Hum. Percept. Perform.* 32:1005–17 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Once+more%3A+Is+beauty+in+the+eye+of+the+beholder%3F+Relative+contributions+of+private+and+shared+taste+to+judgments+of+facial+attractiveness&author=J+Hoenekopp&journal=J.+Exp.+Psychol.+Hum.+Percept.+Perform.&volume=32&pages=1005-17](http://scholar.google.com/scholar_lookup?title=Once+more%3A+Is+beauty+in+the+eye+of+the+beholder%3F+Relative+contributions+of+private+and+shared+taste+to+judgments+of+facial+attractiveness&author=J+Hoenekopp&journal=J.+Exp.+Psychol.+Hum.+Percept.+Perform.&volume=32&pages=1005-17))

Hox JJ. **2010**. *Multilevel Analysis: Techniques and Applications* New York: Routledge [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Multilevel+Analysis%3A+Techniques+and+Applications&author=Hox+JJ&publisher=Routledge](http://scholar.google.com/scholar_lookup?title=Multilevel+Analysis%3A+Techniques+and+Applications&author=Hox+JJ&publisher=Routledge))

Judd CM, McClelland RG, Ryan CS. **2008**. *Data Analysis: A Model Comparison Approach* New York: Routledge [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Data+Analysis%3A+A+Model+Comparison+Approach&author=CM+Judd&author=RG+McClelland&author=CS+Ryan&publication\\_year=2008](http://scholar.google.com/scholar_lookup?title=Data+Analysis%3A+A+Model+Comparison+Approach&author=CM+Judd&author=RG+McClelland&author=CS+Ryan&publication_year=2008))

Judd CM, Westfall J, Kenny DA. **2012**. Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Psychol. Bull.* 138:675–89 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Treating+stimuli+as+a+random+factor+in+social+psychology%3A+a+new+and+comprehensive+solution+to+a+pervasive+but+largely+ignored+problem&author=CM+Judd&author=J+Westfall&author=DA+Kenny&journal=Psychol.+Bull.&volume=138&pages=675-89](http://scholar.google.com/scholar_lookup?title=Treating+stimuli+as+a+random+factor+in+social+psychology%3A+a+new+and+comprehensive+solution+to+a+pervasive+but+largely+ignored+problem&author=CM+Judd&author=J+Westfall&author=DA+Kenny&journal=Psychol.+Bull.&volume=138&pages=675-89))

Kenny DA, Kashy DA, Cook WL. **2006**. *Dyadic Data Analysis*. New York: Guilford Press

Raudenbush SW, Bryk AS. **2002**. *Hierarchical Linear Models: Applications and Data Analysis Methods* Thousand Oaks, CA: Sage [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Hierarchical+Linear+Models%3A+Applications+and+Data+Analysis+Methods&author=SW+Raudenbush&author=AS+Bryk&publication\\_year=2002](http://scholar.google.com/scholar_lookup?title=Hierarchical+Linear+Models%3A+Applications+and+Data+Analysis+Methods&author=SW+Raudenbush&author=AS+Bryk&publication_year=2002))

Satterthwaite FE. **1946**. An approximate distribution of estimates of variance components. *Biom. Bull.* 2:6110–14 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=An+approximate+distribution+of+estimates+of+variance+components&author=FE+Satterthwaite&journal=Biom.+Bull.&volume=2&issue=6&pages=110-14](http://scholar.google.com/scholar_lookup?title=An+approximate+distribution+of+estimates+of+variance+components&author=FE+Satterthwaite&journal=Biom.+Bull.&volume=2&issue=6&pages=110-14))

Smith ER. **2014**. Research design. *Handbook of Research Methods in Social and Personality Psychology* HT Reis, CM Judd 27–48 Cambridge, UK: Cambridge Univ. Press [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Research+design&author=ER+Smith&journal=Handbook+of+Research+Methods+in+Social+and+Personality+Psychology&pages=27-48&publication\\_year=2014](http://scholar.google.com/scholar_lookup?title=Research+design&author=ER+Smith&journal=Handbook+of+Research+Methods+in+Social+and+Personality+Psychology&pages=27-48&publication_year=2014))

Snijders T, Bosker R. **2011**. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* Thousand Oaks, CA: Sage [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Multilevel+Analysis%3A+An+Introduction+to+Basic+and+Advanced+Multilevel+Modeling&author=T+Snijders&author=R+Bosker&publication\\_year=2011](http://scholar.google.com/scholar_lookup?title=Multilevel+Analysis%3A+An+Introduction+to+Basic+and+Advanced+Multilevel+Modeling&author=T+Snijders&author=R+Bosker&publication_year=2011))

Stroup WW. **2012**. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications* New York: CRC Press [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Generalized+Linear+Mixed+Models%3A+Modern+Concepts%2C+Methods+and+Applications&author=WW+Stroup&publication\\_year=2012](http://scholar.google.com/scholar_lookup?title=Generalized+Linear+Mixed+Models%3A+Modern+Concepts%2C+Methods+and+Applications&author=WW+Stroup&publication_year=2012))

Toma C, Corneille O, Yzerbyt V. **2012**. Holding a mirror up to the self: egocentric similarity beliefs underlie social projection in cooperation. *Pers. Soc. Psychol. Bull.* 38:1259–71 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Holding+a+mirror+up+to+the+self%3A+egocentric+similarity+beliefs+underlie+social+projection+in+cooperation&author=C+Toma&author=O+Corneille&author=V+Yzerbyt&journal=Pers.+Soc.+Psychol.+Bull.&volume=38&pages=1259-71](http://scholar.google.com/scholar_lookup?title=Holding+a+mirror+up+to+the+self%3A+egocentric+similarity+beliefs+underlie+social+projection+in+cooperation&author=C+Toma&author=O+Corneille&author=V+Yzerbyt&journal=Pers.+Soc.+Psychol.+Bull.&volume=38&pages=1259-71))

Welch BL. **1947**. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* 34:28–35 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=The+generalization+of+%E2%80%98Student%27s%27+problem+when+several+different+population+variances+are+involved&author=BL+Welch&journal=Biometrika](http://scholar.google.com/scholar_lookup?title=The+generalization+of+%E2%80%98Student%27s%27+problem+when+several+different+population+variances+are+involved&author=BL+Welch&journal=Biometrika))

Westfall J, Judd CM, Kenny DA. **2015**. Replicating studies in which samples of participants respond to samples of stimuli. *Pers. Psychol. Sci.* 10:390–99 [Google Scholar] ([http://scholar.google.com/scholar\\_lookup?title=Replicating+studies+in+which+samples+of+participants+respond+to+samples+of+stimuli&author=J+Westfall&author=CM+Judd&author=DA+Kenny&journal=Pers.+Psychol.+Sci.&volume=10&pages=390-99](http://scholar.google.com/scholar_lookup?title=Replicating+studies+in+which+samples+of+participants+respond+to+samples+of+stimuli&author=J+Westfall&author=CM+Judd&author=DA+Kenny&journal=Pers.+Psychol.+Sci.&volume=10&pages=390-99))

Westfall J, Kenny DA, Judd CM. **2014**. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.*

**title=Statistical+power+and+optimal+design+in+experiments+in+which+samples+of+participants+respond+to+samples+of+stimuli&author=J+Westfall&author=D+45&publication\_year=2014&**

Winer BJ. **1971**. *Statistical Principles in Experimental Design* New York: McGraw-Hill [[Google Scholar](http://scholar.google.com/scholar_lookup?title=Statistical+Principles+in+Experimental+Design&author=Winer+BJ&publication_year=1971)] ([http://scholar.google.com/scholar\\_lookup?title=Statistical+Principles+in+Experimental+Design&author=Winer+BJ&publication\\_year=1971](http://scholar.google.com/scholar_lookup?title=Statistical+Principles+in+Experimental+Design&author=Winer+BJ&publication_year=1971))

**Article Type:** Review Article

## Most Read This Month

### **Self-Compassion: Theory, Method, Research, and Intervention** (</content/journals/10.1146/annurev-psych-032420-031047>)

Kristin D. Neff (</search?value1=Kristin+D.+Neff&option1=author&noRedirect=true>)  
pp. 193–218 (26)

### **Stress and Health: A Review of Psychobiological Processes** (</content/journals/10.1146/annurev-psych-062520-122331>)

Daryl B. O'Connor (</search?value1=Daryl+B.+O%27Connor&option1=author&noRedirect=true>), Julian F. Thayer (</search?value1=Julian+F.+Thayer&option1=author&noRedirect=true>)  
pp. 663–688 (26)

### **The Development of Color Perception and Cognition** (</content/journals/10.1146/annurev-psych-032720-040512>)

John Maule (</search?value1=John+Maule&option1=author&noRedirect=true>), Alice E. Skelton (</search?value1=Alice+E.+Skelton&option1=author&noRedirect=true>) and Anna Franke (</search?value1=Anna+Franke&option1=author&noRedirect=true>)  
pp. 87–111 (25)

### **The Moral Psychology of Artificial Intelligence** (</content/journals/10.1146/annurev-psych-030123-113559>)

Jean-François Bonnefon (</search?value1=Jean-Fran%27ois+Bonnefon&option1=author&noRedirect=true>), Iyad Rahwan (</search?value1=Iyad+Rahwan&option1=author&noRedirect=true>) and Brian Weizenberg (</search?value1=Brian+Weizenberg&option1=author&noRedirect=true>)  
pp. 653–675 (23)

### **Prejudice Reduction: Progress and Challenges** (</content/journals/10.1146/annurev-psych-071620-030619>)

Elizabeth Levy Paluck (</search?value1=Elizabeth+Levy+Paluck&option1=author&noRedirect=true>), Roni Porat (</search?value1=Roni+Porat&option1=author&noRedirect=true>), Chae Eun Lee (</search?value1=Chae+Eun+Lee&option1=author&noRedirect=true>), Donald P. Green (</search?value1=Donald+P.+Green&option1=author&noRedirect=true>) and David A. Asch (</search?value1=David+A.+Asch&option1=author&noRedirect=true>)  
pp. 533–560 (28)

## Most Cited

### **Job Burnout** (</content/journals/10.1146/annurev-psych-52.1.397>)

Christina Maslach (</search?value1=Christina+Maslach&option1=author&noRedirect=true>), Wilmar B. Schaufeli (</search?value1=Wilmar+B.+Schaufeli&option1=author&noRedirect=true>) and Michael P. Leiter (</search?value1=Michael+P.+Leiter&option1=author&noRedirect=true>)  
Vol. 52 (2001), pp. 397–422

### **Executive Functions** (</content/journals/10.1146/annurev-psych-113011-143750>)

Adele Diamond (</search?value1=Adele+Diamond&option1=author&noRedirect=true>)  
Vol. 64 (2013), pp. 135–168

### **Social Cognitive Theory: An Agentic Perspective** (</content/journals/10.1146/annurev-psych-52.1.1>)

Albert Bandura (</search?value1=Albert+Bandura&option1=author&noRedirect=true>)  
Vol. 52 (2001), pp. 1–26

### **On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being** (</content/journals/10.1146/annurev-psych-52.1.141>)

Richard M. Ryan (</search?value1=Richard+M.+Ryan&option1=author&noRedirect=true>), Edward L. Deci (</search?value1=Edward+L.+Deci&option1=author&noRedirect=true>) and Amy C. Ryan (</search?value1=Amy+C.+Ryan&option1=author&noRedirect=true>)  
Vol. 52 (2001), pp. 141–166

### **Sources of Method Bias in Social Science Research and Recommendations on How to Control It** (</content/journals/10.1146/annurev-psych-120710-100452>)

Philip M. Podsakoff (</search?value1=Philip+M.+Podsakoff&option1=author&noRedirect=true>), Scott B. MacKenzie (</search?value1=Scott+B.+MacKenzie&option1=author&noRedirect=true>) and Nathan A. Podsakoff (</search?value1=Nathan+A.+Podsakoff&option1=author&noRedirect=true>)  
Vol. 63 (2012), pp. 539–569

**+** More

#### **About Annual Reviews:**

[What We Do](/about/what-we-do)  
(</about/what-we-do>)

[Press and News](/about/press-center)  
(</about/press-center>)

#### **Discover Content:**

[Journals A-Z](/content/publications) (</content/publications>)

[Impact Factor Rankings](/about/impact-factors) (</about/impact-factors>)

[Publication Dates](/journal/pubdates) (</journal/pubdates>)

[Online Events](/page/events) (</page/events>)

#### **Libraries and Institutions:**

[Subscribe to Open \(S2O\)](/page/librarians/librarian-resource-page) ([/S2O](/page/librarians/librarian-resource-page))

[Librarian Resource Center](/page/librarians/librarian-resource-page)  
(</page/librarians/librarian-resource-page>)

[Institutional Administration Dashboard](#)

#### **Author Resources**

[Article Preparation](#)  
[Submission](#)

</page/authors/get-information>)

[/about/press-center/](#)  
[Careers](#)  
[\(/page/about/careers-at-annual-reviews\)](#)  
[Contact Us](#)  
[\(/page/about/contact-us\)](#)  
[FAQ](#)  
[\(/page/about/faq\)](#)  
[Help \(/help/main\)](#)

[Article Collections \(/page/collectionarchive\)](#)  
[Knowable Magazine \(https://knowablemagazine.org/\)](#)  
[\(https://annurev.publisher.ingentaconnect.com/content/annurev/tca/\)Against the Grain \(https://www.charleston-hub.com/about/about-against-the-grain/\)](#)

[\(https://www.annualreviews.org/admin\)](#)  
[Institutional Pricing](#)  
[\(/page/subscriptions/instchoice\)](#)  
[\(/action/showInstitutionUsageReport\)Against the Grain \(https://www.charleston-hub.com/about/about-against-the-grain/\)](#)

[Editorial Principles](#)  
[Policies](#)  
[\(/page/authors/editing-policies\)](#)  
[Contact Us](#)  
[\(/page/authors/contact-us\)](#)  
[Copyright and Permissions](#)  
[\(/page/about/copyright-and-permissions\)](#)  
[Article Proposals](#)  
[\(/page/authors/author-instructions/unsolicited-manuscripts\)](#)

in

[X](#) [@](#) [\(https://www.linkedin.com/company/annual-reviews\)](#) [\(https://www.facebook.com/annualreviews\)](#) [\(https://www.instagram.com/annualreviews\)](#) [\(https://www.youtube.com/channel/UC...\)](#)