CrossMark

On the ambiguity of interaction and nonlinear main effects in a regime of dependent covariates

Hannes Matuschek[1] · Reinhold Kliegl[2]

Abstract The analysis of large experimental datasets frequently reveals significant interactions that are difficult to interpret within the theoretical framework guiding the research. Some of these interactions actually arise from the presence of unspecified nonlinear main effects and statistically dependent covariates in the statistical model. Importantly, such nonlinear main effects may be compatible (or, at least, not incompatible) with the current theoretical framework. In the present literature, this issue has only been studied in terms of correlated (linearly dependent) covariates. Here we generalize to nonlinear main effects (i.e., main effects of arbitrary shape) and dependent covariates. We propose a novel nonparametric method to test for ambiguous interactions where present parametric methods fail. We illustrate the method with a set of simulations and with reanalyses (a) of effects of parental education on their children's educational expectations and (b) of effects of word properties on fixation locations during reading of natural sentences, specifically of effects of length and morphological complexity of the word to be fixated next. The resolution of such ambiguities facilitates theoretical progress.

Keywords Interaction effects · Mixed models · Additive mixed models · Regression splines · Non-parametric curve estimation

✉ Hannes Matuschek
hannes.matuschek@uni-potsdam.de

[1] Focus Area for Dynamics of Complex Systems and Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24, 14476 Potsdam, Germany

[2] Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24, 14476 Potsdam, Germany

Introduction

Psychological processes are complex and theoretical advances in an understanding of their dynamics are linked with dissociations of their effects in manifest behavior. Frequently, such dissociations are established with interactions between main effects. Here we show that sometimes such interactions may be due to nonlinear main effects and that it may not be possible to distinguish between alternative explanations. This problem arises in many areas across the entire spectrum of psychological research.

In general, we describe a response obtained in a study or an experiment as a function of one or more covariates on which the response may depend.[1] The functional dependency of the response variable on the set of covariates is described in terms of a statistical model that is fitted to observations. We have a wide variety of statistical models at hand to analyze the observations and to estimate the effects of the covariates on the response. In this article, we use analysis of variance (ANOVA) of factorial covariates, linear models (LM), and linear mixed models (LMM) to demonstrate a little known ambiguity between nonlinear main effects and interaction effects in these statistical models. We also present a simple two-step procedure to detect this possible ambiguity.

The ambiguity leads to effects similar to those observed in *suppressor constellations* (e.g., Lewis, & Escobar, 1986; Tzelgov, & Henik, 1991; Friedman, & Wall, 2005). In a suppressor constellation, a covariate (e.g., *z*) that is actually

---

[1]In statistical modeling the response is usually called the *dependent variable* and the covariates are *independent variables*. In this article we examine effects in the presence of statistically dependent *independent variables*. To avoid confusion, we will refer to all *independent variables* as covariates and to the *dependent variable* as the response variable throughout this article.

independent of the response (e.g., *y*) but correlates with a second covariate (e.g., *x*), can improve the fit of a model and may therefore be considered a significant effect. In the following, we demonstrate with a reanalysis of effects of parental education on their children's educational expectations (Ganzach, 1997) and a simple artificial example, that similar effects can be observed between interaction effects and nonlinear main effects if two covariates are correlated. In the "Mathematical background" section, we describe the origin of these artifacts in some detail and extend the issue to nonlinear dependencies between covariates. We also show that nonlinear main effects and interaction effects are actually ambiguous if covariates are dependent and that this ambiguity cannot be resolved by a statistical method alone. Rather, the experimental control of these covariates is required to ensure their independence and to resolve the ambiguity. Having established the mathematical context, we propose a simple two-step procedure to test whether there is an ambiguity between nonlinear main effects and interaction effects. In the "Demonstrations" section, we illustrate the effects of this ambiguity and their detection in a simulation with artificial examples and in a reanalysis of fixation locations during reading.

We illustrate the ambiguity between nonlinear main and interaction effects with a brief reanalysis of children's educational expectations as a function of their parents' education (Ganzach, 1997). Educational expectations of children (EE) are operationalized with the number of years they expect to complete; mothers' (ME) and fathers' (FE) education is indicated by their highest grade achieved. The data were taken from the 1979 *National Longitudinal Survey of Youth (NLSY)* (Center for Human Resource Research, 1995). As Ganzach (1997), we used only the data of those 7748 out of 12,686 children, who were living with both parents and whose mother's and father's education were available in the dataset.

The two covariates ME and FE are correlated $cor(ME, FE) \approx 0.67$. Correlated covariates such as ME and FE are quite common in psychological research and are not limited to surveys like the NLSY; they are also found in many experiments where not all covariates are under experimental control.

Following Ganzach (1997), we analyze the data with two LMs.

$$EE = a\,ME + b\,FE + c\,ME \times FE + \epsilon \tag{1}$$

$$EE = a\,ME + d\,ME^2 + b\,FE + e\,FE^2 + c\,ME \\ \times FE + \epsilon \tag{2}$$

The first model (1) describes educational expectation as a linear function of the mother's and father's education as well as their interaction. In the second LM, we also include

Table 1 Parameter estimates, standard deviations, and t-values for all effects of the LM (1, *left panel*) and LM (2, *right panel*)

	Model Eq. 1			Model Eq. 2		
	Estimate	SE	t-value	Estimate	SE	t-value
ME	0.19	0.01	17.3	0.19	0.01	17.3
ME^2	–	–	–	0.018	0.002	7.41
FE	0.160	0.008	19.8	0.168	0.008	20.6
FE^2	–	–	–	0.014	0.002	8.90
$ME \times FE$	0.017	0.001	11.8	− 0.012	0.003	− 3.80

All effects with a t-value larger than 2, $|t| > 2$, are considered significant. The R^2-values for models (1) and (2) were 0.197 and 0.208, respectively; covariates ME and FE were centered

quadratic effects in ME and FE, capturing some of the possible nonlinearity of the child's educational expectation as a function of mother's or father's educational status.

Table 1 summarizes the LM analyses. For the first LM (1), one finds the intuitive outcome that the child's educational expectation increases with parents' education (positive linear effects of ME and FE) as well as a positive interaction effect, indicating overadditive educational expectation if both parents achieved high grades. The second LM (2) not only finds significant positive quadratic effects of the mother's and father's education on the child's expectation but also a negative interaction effect of the parents' education. Ganzach (1997) concluded that the hypothesized negative (i.e., underadditive) interaction effect was masked by an artifact of the inadequate modeling of the main effects (here linear instead of quadratic) and the correlation of the covariates. We will show, however, that even the negative interaction effect found by Ganzach turns out to be ambiguous as it vanishes if one allows for a flexible description of main effects.

To understand the origin of this ambiguity, one may consider a very simple artificial example: Assume we observed 5000 responses y_i, $i = 1, ..., 5000$ (e.g., response times) in an experiment[2] and also measured two covariates x and z. These covariates reflect the expected processing difficulty of items presented to the subject. Next, let us assume that the covariates x and z are correlated, for example, let $z_i = \frac{x_i}{2} + \frac{u_i}{2}$, where x_i and the unobserved covariate u_i are independent and uniformly distributed in the interval $[-1, 1]$, that is $x \sim U(-1, 1)$, $u \sim U(-1, 1)$. Consequently x and z are correlated with $cor(x, z) \approx 0.70$. Finally, let us assume that the responses were generated by a very simple

---

[2] A large sample ensures a reliable detection of the artifacts. A much smaller sample, however, already shows a strong effect on the type I error rate and power of interaction effects.

Table 2 Summary of ANOVA and LM fit

ANOVA			LM		
Factor	Mean Sq.	F-value	Term	Estimate	t-value
X	0.44	0.421	Main effect: a	0.0164	0.47
Z	0.01	0.006	Main effect: b	-0.0392	-0.79
$X \times Z$	71.79	69.03	Interaction: c	0.955	15.02

There is a strong interaction effect between the factors X and Z (ANOVA; *left panel*) as well as between the continuous covariates x and z (LM; *right panel*)

but nonlinear process that only depends on x but not on z, e.g.,

$$y_i = x_i^2 + \epsilon_i \,,$$

where $\epsilon \sim \mathcal{N}(0, 1)$ is some Gaussian noise. This quadratic relation between the covariate x and the response y implies that they are actually uncorrelated.

The simplest way to analyze such an experiment might be an ANOVA. In this case, the continuous covariates are turned into so-called factors. For example, one may turn the continuous covariate x into a binary variable X with a *negative* level if $x < 0$ and a *positive* level if $x \geq 0$. Analogously, the continuous covariate z is turned into the binary variable Z. As the generating process $y = x^2 + \epsilon$ is quadratic in x, one may not expect any effect of the factor X on y as the response y only depends on the absolute value of x. Therefore, the expectation value of y given $x < 0$ is the same as the expectation value of y given $x \geq 0$. As the generating process does not depend on z at all, we may expect to find neither an effect of Z nor an interaction effect of X and Z on y.

The left panel of Table 2 shows a possible outcome of an analysis of the artificial data. As expected, we do not find strong effects in X and Z. Surprisingly, although the generating process does not depend on the covariate z at all, we find a strong interaction effect (see also left panel of Fig. 1). This interaction effect can be explained by the nonlinear nature of the generating process and the correlation between the covariates x and z: First, as x and z are correlated, it follows that the probability of x and z having the same sign is increased if $|x|$ is large while it is reduced if $|x|$ is small. This implies that $|x|$ is likely to be small if the levels of the factors X and Z differ (i.e., x and z have different signs) and that $|x|$ is likely to be large if the levels match (i.e., x and z have the same sign). Second, as y depends on x^2, it follows that the expectation value of y will be small if $|x|$ is small and large if $|x|$ is large. Consequently, the expectation value of y is increased if the levels of the factors X and Z match, while it is reduced if the levels differ. This is then visible

as a typical interaction effect between the factors in the left panel of Fig. 1.

A similar artifact can be found when analyzing the same data with the linear model (LM)

$$y_i = a\, x_i + b\, z_i + c\, x_i\, z_i + \epsilon_i \text{ where } \epsilon \sim \mathcal{N}(0, \sigma^2)\,, \quad (3)$$

where the coefficients a, b, c and the variance parameter σ^2 are obtained by a fit of the LM to the data. This LM describes the responses y in terms of a sum of linear functions in the continuous covariates x and z ($a\, x_i$ and $b\, z_i$, respectively) and an interaction effect ($c\, x_i\, z_i$). Again, one may expect to find no linear effect of x on the response y as the generating process is quadratic. One may also expect to find neither a linear effect in z nor an interaction effect as the generating process does not depend on z.

The right panel in Table 2 summarizes the results of the LM analysis applied to the same artificial data as used in the ANOVA example. Again, an unexpected strong interaction effect between x and z can be found. As in the ANOVA example, the interaction effect can be explained by the correlation of the covariates x and z and the nonlinear dependency of y on x: If $|x|$ is large, the expectation value of y is increased, as well as the likelihood that x and z have the same sign. Consequently, the product $x \cdot z$ is likely to be positive. If $|x|$ is small, the expectation value of y is reduced, as well as the likelihood that x and z have the same sign. Consequently, the product $x \cdot y$ is more likely to be negative. Therefore, the product $x \cdot y$ is able to capture at least some part of the quadratic effect of x on y in the generating process (see also right panel in Fig. 1). Please note that this artifact is independent of the well-known interpretation issues concerning the scaling and centering of covariates in interaction effects.

This little-known issue of interaction effects of dependent covariates was discussed in Lubinski and Humphreys (1990), Cortina (1993), and Ganzach (1997) for linearly dependent (correlated) covariates and quadratic main effects. In the following section, we generalize this discussion to arbitrarily dependent covariates[3] and general nonlinear main effects. We show that under a regime of dependent covariates, there exists in fact an ambiguity between interaction and main effects that cannot be resolved only with recourse to data. Then we introduce a new two-step method to identify such ambiguous interaction effects, followed by a brief simulation study to demonstrate the procedure. Finally, we apply the method to reanalyses of children's educational expectations and fixation locations during reading.

---

[3]Even if the covariates were uncorrelated, there may still be nonlinear rather than linear dependencies between them. Thus, even if covariates are uncorrelated, it is not ensured that these covariates are statistically independent (see Anscombe's quartet for a beautiful illustration, Anscombe, 1973).
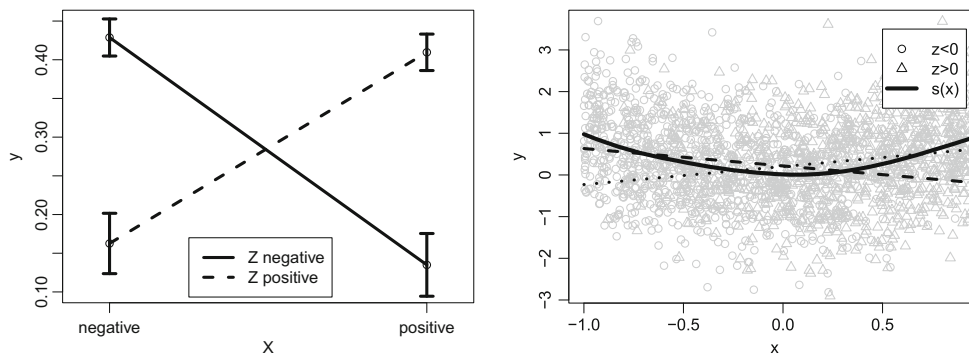
Fig. 1 Visualization of the interaction effect. A typical *crossover* interaction is clearly visible for both the ANOVA (left panel) and the LM (right panel)

Mathematical background

In the following, we discuss a possible origin of ambiguous interaction effects that may be found in the analysis of empirical data. For the sake of simplicity, we will restrict ourselves to simple linear models with polynomial main effects (i.e., x, x^2, etc.) and interaction effects between two mutually dependent covariates (i.e., $x \times z$). The matter, however, generalizes easily to arbitrary nonlinear main effects and higher-order interaction effects in more than two variables, as discussed within this section and demonstrated in the "Simulations" section below. Moreover, the restriction on mere LMs instead of linear mixed models (LMMs, e.g., Pinheiro, & Bates, 2000 also introduced below) eases the discussion of the mathematical background, as spurious fixed effects cannot emerge from neglecting significant random effects (e.g., Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017; Baayen, Vasishht, & Kliegl, 2017).

As in the introduction, we assume that N observations of a response variable y_i, $i = 1, ..., N$ together with some covariates x_i and z_i are obtained in an experiment. Further, we assume that z_i depends linearly on x_i as $z_i = w_x x_i + w_u u_i$, where u_i is a *hidden* covariate that is independent of x_i and $w_x \neq 0$, $w_u \neq 0$. This implies that the covariate z consists of a part that can be explained in terms of the covariate x and a part that is independent of x. Consequently, they are correlated $\text{cor}(x, z) \neq 0$.

Now, assuming that the simple nonlinear process

$$y = x^2 + \epsilon \,, \tag{4}$$

generates the responses y_i, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ are independent and identical distributed (i.i.d.). Given the observations y_i, $i = 1, ..., N$ from this generating process, along with the associated covariates x_i and z_i (please note that the generating process does not depend on z at all), we then may try to explain the observations y_i by means of the linear model

$$y_i = a \cdot x_i + b \cdot x_i z_i + c \cdot z_i + \epsilon_i \,, \tag{5}$$

including an interaction-effect term $x \times z$.

As z depends linearly on x, the model (5) can be expanded to

$$y_i = a \cdot x_i + b \, x_i \underbrace{(w_x \, x_i + w_u \, u_i)}_{=z_i} + c \underbrace{(w_x \, x_i + w_u \, u_i)}_{=z_i} + \epsilon_i$$

$$= a \cdot x_i + b \, w_x \cdot x_i^2 + b \, w_u \cdot x_i \, u_i + c \, w_x \cdot x_i$$
$$+ c \, w_u \cdot u_i + \epsilon_i \,.$$

The expanded model now explicitly contains a quadratic term in x, $b \, w_x \cdot x^2$, which was *hidden* in the interaction effect term (5). This implies that the interaction effect term $x \times z$ of the model (5) contains a part that is implicitly quadratic in x. Therefore, given a sufficiently large sample size, the fit of the model (5) will report a significant interaction effect between the covariates x and z (described by b in Eq. 5), although the generating process (4) neither includes any interactions between these covariates nor depends explicitly on the covariate z at all. This spurious interaction effect originates from the improper description of the main effect of x in the model (5) as a linear one, although the *true* main effect of x on y is quadratic (4).

Unfortunately, the reverse is also true. The quadratic main effect in x may become significant under the following conditions: if the generating process contains an interaction effect between the dependent covariates x and z, such as

$$y = x \cdot z + \epsilon \,, \tag{6}$$

and responses from this process are described by quadratic terms in x and z, but without an explicit interaction effect, such as

$$y_i = a \cdot x_i^2 + b \cdot z_i^2 + \epsilon_i \,. \tag{7}$$

This spurious nonlinear main effect originates from the linear dependence between the two covariates. As $z = w_x x + w_u u$, the generating process (6) is then equivalent to

$$y = x \cdot \underbrace{(w_x \, x + w_u \, u)}_{=z} + \epsilon$$

$$= w_x \, \mathbf{x}^2 + w_u \cdot x \, u + \epsilon \,,$$

which implicitly contains a quadratic term in x (in bold above), that was once again *hidden* in the interaction-effect term of the generating process. Although the generating process (6) is a simple interaction between x and z, the model (7) may report a significant quadratic main effect in x. This implies that in the presence of dependent covariates, an inadequate model for some given observations may result in an ambiguity between interaction and nonlinear main effects.

Of course, if the model matches the generating process, the expected values of the model parameters will match those of the generating process and hence allow for a reliable inference about the existence of interaction or nonlinear main effects. For the analysis of empirical data, however, the generating process is usually unknown. As cognitive processes are complex, one may even assume that no model matches the generating process exactly (Box, 1979). This is usually not a problem, as a chosen model may approximate the generating process sufficiently well, at least in the partial effects of theoretical interest. The presence of mutually dependent covariates, however, implies an ambiguity between the interaction effects and nonlinear main effects, which cannot be resolved. As interaction effects are typically interpreted in a completely different way than nonlinear main effects, the ambiguity must be taken into account or at least discussed in the report.

For simple cases like those above, where covariates are only linear dependent (i.e., correlated), the ambiguity between nonlinear main and interaction effects can indeed be resolved by including polynomial main effects up to the same order as the sum of the polynomial degrees of all interaction effects in the model. For example, a model with one interaction effect $x \times z$ should also include linear and quadratic main effects in x and z (e.g. Lubinski, & Humphreys, 1990; Cortina, 1993; Ganzach, 1997). This solution, however, is only valid if one can assume that the dependency between the covariates x and z is linear. For covariates derived from some natural setting, it is not ensured that the assumption of a linear dependency is justified. If this dependency is nonlinear, the ambiguity cannot be resolved, and an interaction effect may then also be explained by a polynomial main effect of a higher order than 2 and vice-versa (see "Demonstrations" section below).

So far, we have described an effect where spurious interaction effects may appear in the presence of linear dependent (correlated) covariates and when nonlinear main effects are inappropriately described in a statistical model (e.g., by linear ones).

If the dependency between the covariates (here x and z) is nonlinear (e.g., $z = w_x \cdot x^2 + w_u \cdot u$), where x and u are independent, the description of the data sampled from a generating process containing a cubic term, for example,

$$y = x^3 + \epsilon \,,$$

using a model with an interaction effect between the covariates x and z, for example,

$$y_i = a \cdot x_i^2 + b \cdot x_i\, z_i + c \cdot z_i^2 + \epsilon_i \,,$$

leads to a significant interaction effect between x and z, as the interaction effect term $x_i\, z_i = w_x \cdot x_i^3 + w_u\, x_i\, u_i$ implicitly contains a cubic term in x.

According to Lubinski and Humphreys (1990), Cortina (1993), and Ganzach (1997), the ambiguity between nonlinear main effects and interaction effects should be resolved for the latter model. In fact, this is only the case if the dependency of the covariates x and z is linear. If the dependency is nonlinear, the ambiguity between main and interaction effects reappears.

Alternatively, the dependency between the covariates x and z can be *removed* by residualizing a covariate, for example z, with respect to the other covariates (e.g., x). The aim of this method is to obtain a new covariate, $\tilde{z}$, which is independent of all other covariates (e.g., Wurm, & Fisicaro, 2014). Residualization, however, is usually a parametric approach. That is, residualizing z as $z = a\, x + b\, x^2 + \tilde{z}$ will ensure that x and $\tilde{z}$ are independent with respect to linear and quadratic terms. Any higher-order dependency between the covariates x and z will remain and may still imply an ambiguity of interaction effects between x and the residualized covariate $\tilde{z}$ and nonlinear main effects.

Although we restrict ourselves to linear models in the discussion of this phenomenon,[4] the problem also appears in generalized linear (mixed) models (e.g., Demidenko, 2013) and even in simple ANOVAs by splitting a continuous variable into two categories (e.g., $x < 0$ and $x \geq 0$), as demonstrated above.[5] The effect will appear as a significant interaction effect of the associated categorical variables if the generating model contains a quadratic term (e.g., $y = x^2 + \epsilon$) and the covariates x and z are linearly dependent.

Detecting ambiguous main and interaction effects

In the following, we introduce a simple two-step approach to detect a possible ambiguity between nonlinear main effects

---

[4]Please note that a similar ambiguity between the main and some interaction effects may arise in cases where the generating process is an unknown nonlinear but monotonic mapping of a truly additive process. For example, $y_i = f(a\, x_i + b\, z_i + c\, x_i\, z_i) + \epsilon_i$ (Garcia-Marques, Garcia-Marques, & Brauer, 2014). As the function $f(\cdot)$ is unknown, the ambiguity between the main and interaction effects cannot be resolved even if the covariates x and z are independent. This type of interpretation issue is different from the one discussed here and it can neither be resolved by ensuring the independence of the covariates nor identified by a non-parametric description of the main-effect terms.

[5]Of course, covariates should definitely not be converted to categorical variables to begin with (e.g., McClelland, Lynch, Irwin, Spiller, & Fitzsimons, 2015).

and interaction effects. In general, neither the exact main-effect functions nor the exact dependencies between the covariates are known, hence an appropriate description of the main effects becomes increasingly difficult, especially when relatively large samples are involved. In these cases, even small discrepancies between the parametric model and the *true* generating process, as well as weak dependencies between the covariates, may lead to highly significant interaction effects. This issue reflects the general limitation of parametric approaches to describe a functional dependency of a response variable on covariates.

Therefore, we need a non-parametric and adaptive method for the description of the main effects that allows for an increasing flexibility in the description of main effects as more and more data become available. Such an approach avoids the problem that even a relatively small mismatch between the parametric model and the *true* main effects yields a significant result, because the method adapts to the increased sensitivity of the statistics with increasing sample sizes.

Splines are a versatile tool. They allow for exactly this adaptive and nonparametric description of the main-effect functions (Silverman, 1985). They are generically smooth or at least continuous functions in one or more variables and are obtained by searching for a trade-off between the goodness-of-fit to the data and the *wiggliness* (complexity) of the function. In most cases, the so-called *thin-plate regression spline* can be used. A spline is a function $s(x)$, given the data y_i at $x_i, i = 1, ..., N$, that solves the optimization problem

$$s(x) = \underset{f(x)}{\operatorname{argmin}} \left(\sum_{i=1}^{N} |y_i - f(x_i)|^2 + \lambda \int_{-\infty}^{\infty} \left| \frac{\partial^2 f(x)}{\partial x^2} \right|^2 dx \right), \tag{8}$$

where λ is the parameter that determines the trade-off between the goodness-of-fit ($\sum_i |y_i - f(x_i)|^2$) and the *wiggliness* of the function ($\int \left| \partial_x^2 f \right|^2 dx$).

The close relation between spline estimates and LMs (Kimeldorf and Wahba, 1970; Wahba, 1990), not only led to the unification of these two tools into one, the additive model (AM, e.g., Wood, 2006), but also allows us to determine the trade-off between the goodness-of-fit and *wiggliness* (λ) by means of maximum likelihood (e.g., Wood, 2006). In contrast to an LM, an AM allows for the description of main effects by means of arbitrary spline functions instead of parametric terms like polynomials. Recent advances in the inference methods of AMs (Wood, 2003; Wood, Scheipl, & Faraway, 2012; Bates, Mächler, Bolker, & Walker, 2015) have made it possible to apply these techniques even to large datasets with many covariates (Bates, Mächler, Bolker, & Walker, 2014; Wood, 2014). Readers interested in non-parametric spline fits may want

to consult Wahba (1990) for mathematical details or Wood (2006) and Appendix for applied perspectives.

Of course, the flexibility of AMs does not come without a cost. Although spline functions in AMs can be considered *nonparametric main effects*, they are actually penalized in order to make the spline regression problem (8, above) uniquely identifiable and to avoid over-fitting the splines to the data. This penalty towards smoother functions may introduce a small bias in favor of the completely unpenalized parametric interaction effects of an AM. This implies that, even if the interaction-effect term of the AM, $y_i = a \cdot x_i z_i + s_x(x_i) + s_z(z_i) + \epsilon_i$ can be explained entirely by the nonlinear main-effects splines $s_x(x)$ and $s_z(z)$, the penalty towards smoother functions will introduce a small bias on the estimate of a and hence increase the interaction effect size.

In order to reduce this bias, one may prevent the *competition for variance* between the interaction-effect term and the main-effect splines with the following two-step procedure. First, a pure main effects AM (see Wood, 2006, for an introduction to AM/AMM fits) is fitted to the data:

$$y_i = s_x(x_i) + s_z(z_i) + \epsilon_i . \tag{9}$$

Second, a simple LM containing only the interaction effect term of interest is fitted to the residuals of the first model:

$$\epsilon_i = a \cdot x_i z_i + \epsilon_i' . \tag{10}$$

The second model allows us to determine the size of the interaction effect between the covariates x and z, which cannot be explained by the nonlinear main effects in Eq. 9; it also allows us to test whether there remains a significant interaction effect.

Similar to the case discussed above (6 and 7), an additive main-effects model can also explain at least some part of an interaction effect that is present in the generating model. Therefore, the ambiguity between the nonlinear main- and interaction-effect sizes remains: If nonlinear main effects are not adequately modeled, the type-I error rate of interaction effects might be increased, while the power to detect existing interaction effects might be reduced as they can be described at least partially by nonlinear main effects.

Demonstrations

In this section, we demonstrate the effects of the ambiguity described in the "Mathematical background" section and the ability of our two-step procedure to detect that ambiguity. First, with simple simulations using LMs, we compare the ability to discover ambiguities between previous parametric approaches and our non-parametric two-step procedure using AMs. Second, with two *real-data* examples, we demonstrate the ambiguity between an interaction

effect and nonlinear main effects with analyses of children's educational expectations using LMs and fixation locations during reading of Uighur sentences with linear mixed models (LMMs).

Simulations

In this brief simulation study, we demonstrate some effects of the ambiguity between nonlinear main effects and interaction effects in the presence of mutually dependent covariates (here x and z) and show that our two-step nonparametric approach to detect these ambiguities performs well where current parametric approaches fail.

For each of the 100 simulation iterations, the covariates for the evaluation of the generating processes are sampled as following: First, 1000 independent and identically distributed random values for x and u are sampled uniformly from the interval $[-1, 1]$. Then, the second covariate z is obtained as $z = \frac{x}{3} + \frac{2u}{3}$ for simulations 1–5 and as $z = 4x^2 + u$ for simulations 6 and 7. The first case induces linear-dependent covariates x and z, where the correlation between these covariates is about $\mathrm{cor}(x, z) \approx 0.45$. The second case leads to uncorrelated but nonlinear-dependent covariates.

Table 3 shows a summary of the simulation results. In Simulation 1 (first row in Table 3), the generating process is $y = x^2 + \epsilon$ and the model is $y = a \cdot x + b \cdot z + c \cdot x z + \epsilon$. The simulation results show an average t-value $\bar{t}$ for the interaction effect term $x \times z$ of about $\bar{t} \approx 3.57$, suggesting a highly significant interaction effect, although the generating process neither includes such an interaction effect, nor does it depend on the covariate z at all. As described in the introduction, this spurious interaction effect originates from the linear dependency of covariates x and z and from the inadequate (incorrect) model specification with respect to the *true* nonlinear (here quadratic) main effect in x that is part of the generating process. In this case, the ambiguity between nonlinear main effects and interaction effects results in a severe increase in the type-I error rate.

In Simulation 2 (second row in Table 3), the model properly accounts for this nonlinearity by including a quadratic term in x. In this case, the average t-value is about $\bar{t} \approx 0.19$, which agrees with the absence of a significant interaction effect.

In Simulation 3, the observations y are first described by an AM with two main-effect splines $s_x(x)$ and $s_z(z)$. A second linear model was then used to check if there remains an interaction effect between the covariates in the residuals of the AM. The AM approach allows for arbitrary smooth functions as main effects, in contrast to the linear models in simulations 1 and 2, where a parametric approach with polynomial main effects was used. The results show that no significant interaction effect is found as the AM splines are able to capture the nonlinear main effects.

Table 3 Simulation results for different generating processes and models

Generating process	Model	$\bar{R}^2$-value	$\bar{t}$-value
1 $y = x^2 + \epsilon$	$y = a \cdot x + b \cdot z + c \cdot \mathbf{x}\mathbf{z} + \epsilon$	≈ 0.017	≈ 3.84
2 $y = x^2 + \epsilon$	$y = a \cdot x + b \cdot x^2 + c \cdot \mathbf{x}\mathbf{z} + \epsilon$	≈ 0.084	≈ 0.10
3 $y = x^2 + \epsilon$	$y = s_x(x) + s_z(z) + \epsilon$	≈ 0.086	
	$\epsilon = c \cdot \mathbf{x}\mathbf{z} + \epsilon'$	< 0.001	≈ 0.12
4 $y = x \cdot z + \epsilon$	$y = s_x(x) + s_z(z) + \epsilon$	≈ 0.042	
	$\epsilon = c \cdot \mathbf{x}\mathbf{z} + \epsilon'$	≈ 0.012	≈ 3.46
5 $y = x^3 + \epsilon$	$y = a \cdot x + b \cdot x^2 + c \cdot \mathbf{x}\mathbf{z} + \epsilon$	≈ 0.11	≈ -0.069
6 $y = x^3 + \epsilon$	$y = a \cdot x + b \cdot x^2 + c \cdot \mathbf{x}\mathbf{z} + \epsilon$	≈ 0.12	≈ 4.14
$z = 4x^2 + u$			
7 $y = x^3 + \epsilon$	$y = s_x(x) + s_z(z) + \epsilon$	≈ 0.13	
$z = 4x^2 + u$	$\epsilon = c \cdot \mathbf{x}\mathbf{z} + \epsilon'$	< 0.001	≈ 0.34

The last column shows the approximate mean t-value over 100 iterations for the interaction effects of the model (*in bold*). For each iteration of the simulation, each analyzing model was fitted to a set of 1000 independent samples from the generating process. An interaction effect is considered significant if the average t-value is greater than two, $|\bar{t}| > 2$

In Simulation 4, we demonstrate that a *true* interaction effect between the dependent covariates x and z can still be detected reliably. The second model, fitted to the residuals of the AM, shows a highly significant interaction effect that could not be explained by the smooth main effects in x and z. Please note that the average value of the estimated interaction-effect coefficient a is $\bar{a} \approx 0.45$, although the interaction-effect coefficient of the generating process is $a = 1$. This indicates that at least some part of the *true* interaction effect was explained by the main-effect splines. Consequently, the ambiguity between nonlinear main effects and interaction effects reduces the power to detect interaction effects here.

Simulations 5–7 demonstrate the effects of linear and nonlinear dependencies between the covariates x and z. In Simulation 5, the parametric model correctly states that there is no significant interaction effect, although the cubic main effect in x of the generating process is not described properly by the model. As discussed above, this is due to the linear dependency between the covariates x and z. The parametric model does not contain a term that is able to explain the cubic main effect in x of the generating process. If, however, the dependency between the covariates x and z is nonlinear, as is the case in simulations 6 and 7, the interaction effect in the model becomes significant again (Simulation 6). In this case, the interaction effect contains a cubic contribution in x and hence describes the cubic main effect of the generating process as an interaction effect between x and z. This, in turn, increases the type I error

rate for the interaction effect. Using the AM approach (Simulation 7) solves this issue, as the main-effect splines will adapt to complex shaped functions if sufficient evidence is provided by the data.

Analysis of educational expectations

In this section, we continue with the example from the introduction (Ganzach, 1997). We test whether the negative interaction effect (see Eq. 2) is ambiguous using the AM approach introduced here. First we model the dependency of the child's educational expectation as a sum of arbitrary but smooth nonlinear functions of the parents' education (11) and then we test whether a significant interaction effect between these covariates is still present in the residuals of the AM (12).

$$EE = s(ME) + s(FE) + \epsilon \tag{11}$$

$$\epsilon = ME \times FE + \epsilon' \tag{12}$$

The results of the LM (12) show that there is no significant interaction effect in the model residuals (11) ($M = -0.00198$, $SD = 0.00129$, $t = -1.53$, $R^2 = 0.214$ of AM (11) and $R^2 < 0.001$ of LM (12)). These results suggest that non-linear main effects can account for the negative interaction effect of the correlated parents' education found in the presence of simple polynomial main effects (here modeled with linear and quadratic terms). Thus, there is a strong ambiguity between non-linear main and linear interaction effects.

Analysis of fixation locations during reading

In our research on eye-movement control during reading, visual (e.g., word length) and lexical (e.g., word frequency) variables influence where we look and for how long (e.g., Hohenstein, Matuschek, & Kliegl, 2016). Usually, these and many other variables are correlated and, as far as reading of natural sentences is concerned, not all of them can be controlled in an experiment. For example, word length and frequency are naturally correlated, as shorter words are generally more frequent than longer words. Or more specifically, to foreshadow the example we will use in this section, in some languages, such as in Uighur script, multiple suffixes coding gender, case, or number may be serially tacked onto the root of a noun. Consequently, long words will usually be morphologically more complex (i.e., carry a larger number of suffixes) than short words. Yan, Zhou, Shu, Yusupu, Miao, Krügel, & Kliegl (2014) showed that eye movement programs are not only influenced by the length of the next word (i.e., fixations are usually close to the center of words irrespective of their length; Rayner, 1979), but also its morphological complexity (i.e., fixations are closer to the beginning of words with multiple suffixes). The

theoretical relevance of this result is that programming an eye movement to the next word is not only based on visual information (i.e., the length of the next word, delineated by clearly marked spaces between words) but also by subtle information that requires an analysis of within-word details (i.e., identification of suffixes). Thus, the results suggest that we extract quite a bit of linguistically relevant detail from a word before it is fixated. Obviously, these effects are very small and, moreover, high correlations between variables may dramatically reduce the statistical power to detect them. Matters are even worse if we want to test the interaction between continuous covariates such as word length and morphological complexity (McClelland and Judd, 1993). Examples such as this motivated our research, but, as mentioned above, the dissociation of subtle effects between correlated variables is a pervasive concern in most areas of psychological research.

In this section, we test whether such an significant interaction effect can be explained by nonlinear main effects in a regime of dependent covariates. Specifically, we reanalyze fixation locations measured during reading of Uighur script as function of covariates relating to the fixated word (Yan et al., 2014). Forty-eight undergraduate students from Beijing Normal University, all of them native speakers of Uighur, read 120 Uighur sentences. Each subject read half of the sentences in Uighur and the other half in a Chinese translation. The analysis is based only on eye movements during Uighur reading. The variables of theoretical interest are the length of the words and of their root morphemes, as well as the length and number of suffixes. Word length varied from 2 to 21 letters ($M = 7.5$, $SD = 3.0$). The percentages of words with 0, 1, 2, and more than 2 suffixes were 34, 38, 19, and 9%. The length of the root morphemes varied from 1 to 11 letters ($M = 4.7$, $SD = 1.7$) and the total number of letters in suffixes varied from 1 to 15 ($M = 3.9$, $SD = 2.3$). The data set comprises a total of 13,523 fixations. For further details on the experimental setup and analyses using LMMs we refer to Yan et al. (2014).

Although linear models are frequently used for the analysis of experimental data, it is more adequate to resort to a broader class of models, the linear mixed models (LMMs, e.g., Pinheiro, & Bates, 2000; Bates et al., 2015). By including random effects, LMMs can be seen as a generalization of LMs. The random effects describe the deviance of the response from the ensemble mean (the fixed effects LM part) for a grouping factor as a sample of a common (normal) distribution. For example, a random intercept for each individual allows us to model the deviation of the responses of each subject from the ensemble mean (described by the fixed effects). Frequently, not only individual differences are modeled as random effects but also item-specific differences (e.g., random effects of words and sentences in a reading experiment). Analogous to the extension of LMs to LMMs,

it is possible to extend additive models to additive mixed models (AMMs, e.g., Wood, 2006; Matuschek, Kliegl, & Holschneider, 2015). The following reanalysis uses LMMs and AMMs as statistical models, not LMs and AMs.

The Uighur script is well known for its rich usage of up to five suffixes. As the fixation location x_l within a word depends on the length of the fixated word l_w as well as on the suffix length l_s of that word (Yan et al., 2014), an LMM describing the fixation location x_l will contain these two trivially correlated covariates. The word and suffix lengths are linear dependent, as the length of a word is simply the sum of the root morpheme and the suffix length, $l_w = l_r + l_s$. Indeed, the correlation coefficient between these two covariates is about $\text{cor}(l_w, l_s) \approx 0.7$ in the Uighur sentence corpus (USC, Yan et al., 2014).

Of course, this direct linear dependency between word and suffix lengths can be resolved easily by replacing the word-length covariate l_w with the length of the root morpheme l_r. In terms of cognitive processing, however, word length maps onto visual processing whereas the distinction between the length of the root morpheme and the suffixes requires within-word sublexical processing. Thus, there is a theoretical argument for using word length as a covariate. Moreover, we want to demonstrate spurious interaction effects in a *real-world* example using an obvious linear dependency between two covariates. Replacing the covariate l_w by l_r does not solve the issue of dependent covariates, because the length of the root morpheme is still correlated with the length of the suffix, $\text{cor}(l_r, l_s) \approx 0.46$.

Fixation locations in words also depend strongly on the amplitude of the incoming saccade (a) (McConkie, Kerr, Reddix, & Zola, 1988; Engbert, & Krügel, 2010), that is the distance of the last fixation location from the beginning of the fixated word. Hence, a simple LMM for the analysis of the fixation location x_l could be

$$x_l \sim \left(a + a^2\right) \times l_w \times l_s + (1|Word) + (1|Sentence)$$
$$+ (1|Subject). \tag{13}$$

This LMM (13) is presented in the model notation of the *lme4* R package (Bates et al., 2014; R Core Team, 2014). It describes the fixation location on letter x_l and incorporates as fixed effects all possible interactions between a quadratic polynomial of the incoming saccade amplitude a and the linear effects of word and suffix lengths of the fixated word. To account for between-subject and between-item variability, the LMM also includes random intercepts for the fixated word, sentence, and subject.

Table 4 (left panel) summarizes the results for the fixed effects obtained for the LMM (13). Effects are considered significant (printed in bold) if the absolute value of their t-value is larger than 2. In this brief analysis, we find significant main effects of incoming saccade amplitude (a, linear part), word length (l_w), and suffix length of the

Table 4 Parameter estimates, standard deviations, and t-values for all fixed effects of the LMM (13, left panel) and LM (15, right panel)

	Model (13)			Model (15)		
	Estimate	SE	t-value	Estimate	SE	t-value
a	-103	3.11	$-\mathbf{33.3}$	-0.93	2.79	-0.334
a^2	4.37	3.03	1.44	-1.24	2.82	-0.440
l_w	1.36	0.07	$\mathbf{18.9}$	0.001	0.03	0.019
l_s	-0.09	0.04	$-\mathbf{2.34}$	-0.002	0.01	-0.177
$a \times l_w$	-40.6	4.80	$-\mathbf{8.46}$	-35.0	4.42	$-\mathbf{7.93}$
$a^2 \times l_w$	2.83	4.67	0.61	2.45	4.32	0.567
$a \times l_s$	-0.21	2.46	-0.08	-0.54	2.28	-0.235
$a^2 \times l_s$	1.54	2.44	0.63	1.33	2.28	0.586
$l_w \times l_s$	0.14	0.05	$\mathbf{3.06}$	0.01	0.02	0.379
$a \times l_w \times l_s$	3.64	3.13	1.17	3.21	2.90	1.10
$a^2 \times l_w \times l_s$	-0.658	3.04	-0.22	-1.25	2.82	-0.443

All effects with a t-value larger than 2, $|t| > 2$, are considered significant and printed in bold font. The R^2-values for models (13), (14), and (15) were 0.486, 0.482, and 0.00962, respectively

word (l_s). Additionally, we find two significant interaction effects between incoming saccade amplitude and word length ($a \times l_w$), as well as between word length and suffix length ($l_w \times l_s$).

Word length l_w trivially correlates with suffix length l_s. Under the assumption that the main effect of suffix length l_s on fixation location x_l is nonlinear (please note that the LMM above describes these main effects as linear), the significant interaction effect between word length and suffix length ($l_w \times l_s$) might be ambiguous. As described above, we test these potential nonlinearities in the main effects by first fitting the AMM (14) to the fixation locations, including generic smooth functions in a, l_w and l_s as main effects. Then we test whether the residuals of this AMM still contain significant interaction effects which could not be explained by the nonlinear main effects with a simple LM (15).

$$x_l \sim s_a(a) + s_{l_w}(l_w) + s_{l_s}(l_s)$$
$$+ (1|Word) + (1|Sentence) + (1|Subject) \tag{14}$$
$$\epsilon \sim \left(a + a^2\right) \times l_w \times l_s \tag{15}$$

The right panel of Table 4 summarizes the parameter estimates, standard deviations, and t-values of the fixed effects of the LM (15). Unsurprisingly, the main effects are no longer significant, as they are described by the spline main-effects in the AMM (14). Concerning the interaction between incoming saccade amplitude and word length ($a \times l_w$), we find that only a relatively small part of this effect is explained by the nonlinear main effects, as it is still highly significant and its coefficient drops only slightly from about -40 to about -35 (compare left and right panels in Table 4). The interaction between word and suffix length
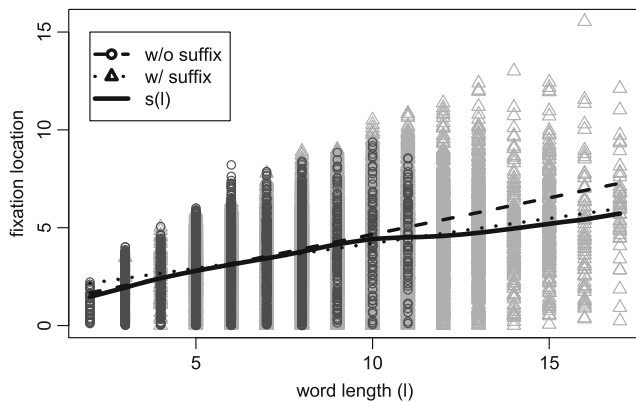
Fig. 2 Scatter plots of fixation location and word length for words with and without suffixes (shown as triangles and circles, respectively) and linear main effects of word length on fixation location for words with and without suffixes (dotted and dashed lines, respectively). The solid line shows the spline main effect of word length on fixation location

$(l_w \times l_s)$, however, is explained completely by the nonlinear main effects of the AMM (14). According to the LM (15), no significant interaction effect remains in the residuals of the AMM (14).

Figure 2 visualizes the ambiguity between a possible nonlinear main effect of word length l_w on fixation location x_l and a possible interaction effect between word length and suffix length l_s. This figure shows the scatter plots and linear relationships of fixation locations and word lengths for words without a suffix (circles / dashed line) and words with a suffix (triangles / dotted line). The solid line shows a spline fit describing a nonlinear relationship between word length and fixation location. The scatter plots show that words without suffixes are generally shorter than words with suffixes and the linear trends show that the slopes differ between words with and without suffixes. The latter might be interpreted as an interaction effect between word length and a factor *suffix*. The spline-fit, however, can explain the same effect by allowing for a nonlinear main effect in word length. In fact, the spline appears to follow the linear trend for words without suffixes up to word length $l_w = 11$, and to follow the linear trend of words with suffixes for $l_w > 11$.

These results, however, do not necessarily imply that the interaction effect does not exist in the first place. The test only informs us that there is a strong ambiguity between the nonlinear main- and interaction-effect terms due to dependencies between these covariates.

Discussion

We have shown, theoretically, with simulations, and with reanalyses of children's educational expectations (Ganzach, 1997) and of fixation locations in a reading experiment (Yan

et al., 2014), that an ambiguity between nonlinear main effects and interaction effects may exist in regimes of dependent covariates. The applications illustrate that this issue shows up in social-science studies and cognitive-science experiments. It may appear in a wide range of analysis tools, such as (G)LMMs, LMs, and even simple ANOVAs. This ambiguity leads to effects similar to what are known as suppressor constellations. However, it extends beyond suppressor constellations in two ways: First, in contrast to suppressor constellations, it is not a variable but rather an interaction effect or a nonlinear main effect that acts as the *suppressor*. Second, as we extend the dependency between covariates from linear dependent (correlated) to arbitrarily nonlinear dependent covariates, the ambiguity between nonlinear main effects and interaction effects cannot be resolved by statistical analyses alone. To this end, it is impossible to determine which term is the actual *suppressor*. It is therefore crucial to detect whether there is an ambiguity or not.

We propose a novel method to test for this ambiguity, using AMs that allow for a description of nonlinear main effects by means of generic smooth functions (splines). These splines are able to *adapt* to the increase of information often provided by increasing sample sizes; they provide an appropriate description of arbitrarily complex, but smooth functions. This adaptive behavior cannot be achieved by resorting to parametric approaches, such as polynomials of a chosen degree.

Applying this novel method to data from a free-reading experiment of Uighur script (Yan et al., 2014), we found that one of two significant interaction effects could be explained—one could say: completely—as being due to nonlinear main effects. Regrettably, neither this test nor a non-parametric approach to the residualization of covariates provides proof of whether or not an interaction effect is spurious. However, the test is able to detect ambiguities between possible nonlinear main effects and interaction effects. If the ambiguity is strong, meaning in cases where an interaction effect can be explained to a large degree (completely) by nonlinear main effects, then, if at all possible, an additional experiment should be performed that controls for the mutually dependent covariates involved in the ambiguous interactions. Only experimental control of the covariates will resolve the ambiguity. To this end, we are not yet able to provide a definite threshold for what constitutes a *strong* ambiguity. This is due to the fact that a change of an interaction effect, which was significant in an LMM with linear main effects and non-significant in an AMM assuming non-linear main effects, is itself not necessarily significant. Moreover, it is difficult to state the significance of a change in the interaction effect size, as both effect estimates are based on the same data. Hence, a reliable test statistic to determine whether a significant ambiguity is present will need to be developed.

In our opinion, it is crucial to test whether interaction effects found in an analysis using LMs, LMMs, GLMMs, or even ANOVAs involving dependent covariates can be explained by assuming nonlinear main effects. If this is the case, there are potentially strong implications for their interpretation. For example, the type I error rate of interaction effects might be increased severely if nonlinear main effects are not modeled adequately. On the other hand, the power of existing interaction effects might be reduced, as they can at least partially be explained by nonlinear main effects. All else being equal, if an interaction effect can be explained completely (compare left and right panels in Table 4) by nonlinear main effects, we suggest to consider the interaction effect to be ambiguous, as it can be eliminated from the model by allowing for nonlinear main effects. This represents the usual appeal to parsimony, as nonlinear main effects are usually in line with theoretical models (as long as they are monotonic), while complex interactions usually require much more complex theoretical models.

Once we move from quadratic polynomial main effects to splines to account for interactions between main effects, as in the reanalysis of children's educational expectations (Ganzach, 1997), one could argue that we are also softening assumptions about monotonic effects; splines are usually wiggly, not strictly monotonic. Including an interaction term might be more parsimonious than the spline-based main effects. However, there are two counterarguments. First, strictly speaking: a quadratic main effect only appears as monotonic as long as it is restricted to values below or above the minimum or maximum of the function. In this respect, linear and quadratic main effects just as main effects based on splines both represent only an approximate description of a functional relation. Second, if wiggles in main effects replicate, but are small relative to a large monotone declining or increasing trend associated with a main effect, they most likely reflect influences of covariates not yet in the model. Such systematic violations of a monotonic trend may help with the identification of these moderating covariates. Their inclusion will significantly reduce the "model wiggliness". Our preference, therefore, is that such an ambiguity be resolved with reference to the current state of theory. The specification of a statistical model should be motivated from a theoretical perspective–rather than the hope that the statistical model (i.e., the data analysis) will deliver a theoretical perspective. Thus, we hope to raise awareness of the interpretational options and their associated constraints. For example, from the current theoretical perspective, the interpretation of the results as an interaction may be preferred. However, future theoretical developments may favor the nonlinear main-effect interpretation. The bottom line is that, ideally, multivariate statistics should be in the service of theory-guided research and not vice versa.

Appendix: Splines and additive mixed models

Splines are versatile tools for non-parametric modeling of arbitrary functions. They approximate some unknown function $g(x)$ of which only a finite number of noisy observations y_i at locations x_i are given, i.e. $y_i = g(x_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2) \, i.i.d.$. Of cause, there are infinite many different functions that would describe the noisy observations perfectly. Hence one assumes a certain penalty $J(f)$ on the function to describe the data. Frequently, the so called *wiggliness* is used as a penalty. That is, the integral over the squared second derivative of the function

$$J(f) = \int \left| \frac{\partial^2 f}{\partial x^2} \right|^2 \, dx \, .$$

This penalty can be considered as a complexity measure. We search for a function f to describe the observed data that minimizes the penalty $J(f)$, which is the least *wiggly* or *bumpy* function. Although the introduction of the penalty $J(f)$ ensures that there is only one function that describes the observations, it may still overfit the data as there exists a single function $f(x)$ that will reproduce the observed data exactly, i.e. $y_i = f(x_i)$. Hence one may search for a trade-off between goodness-of-fit of the function to the data and the penalty $J(f)$ (wiggliness)

$$\underset{f}{\text{argmin}} \, \underbrace{\|y_i - f(x_i)\|^2}_{\text{goodness-of-fit}} + \lambda^2 \, \underbrace{J(f)}_{\text{penalty}} \, . \qquad (16)$$

For this optimization problem, the parameter λ determines the trade-off between the goodness-of-fit and the penalty. If λ is large, the wiggliness-penalty is strong and a very smooth function will be selected, while for a small λ the goodness-of-fit term dominates and a more complex function will be selected. In fact, there is a simple physical analogy to splines and the goodness-of-fit/penalty trade-off.

Consider a thin elastic rod suspended with springs at some locations in the plane (see Fig. 3). These locations represent the observations. The amount by which the springs are extended, represents the goodness of fit and the bending of the rod the *wigglyness* of the solution $f(x)$. The aforementioned trade-off is determined by the ratio of the stiffness of the springs (Hooks constant k) and the stiffness of the rod (Young's modulus E). Hence, the stiffer the rod compared to the springs, the smoother the solution, while
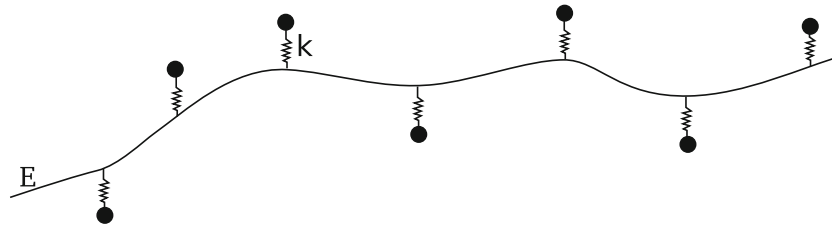
Fig. 3 Physical analogy to a spline: an elastic rod suspended in springs. The solution is determined by the stiffness of the springs and rod, as the minimization of the total mechanical energy stored in the bended rod and extended springs

stiffer springs will pull the rod closer to the fixation points (observations) leading to a more wiggly result.

Neither the mathematical formulation of the problem (16) nor its physical analog (Fig. 3) allows for a determination of λ as the best trade-off between goodness-of-fit and wigglyness. To achieve this, one has to transform the function-approximation problem above (16) to a statistical regression problem. Hence a theory is required that connects the world of function approximation with the world of statics. The theory of *reproducing kernel Hilbert spaces* (RKHS) does exactly that.

A RKHS theorem states that the solution of the optimization problem above (16) lays within a RKHS uniquely determined by a single function, the reproducing kernel $R(x, x')$, where the optimal solution is given by a finite linear combination of these kernels $f(x) = \sum_i \alpha_i R(x, x_i)$ (e.g., Matuschek, 2016; Matuschek et al., 2015; Berlinet, & Thomas-Agnan, 2004; Wahba, 1990; Kimeldorf, & Wahba, 1970). These kernel functions $R(\cdot, \cdot)$ have another helpful property, they are positive definite functions. This means, they are so-called precision functions.[6] Moreover, the precision matrix $\Gamma_{i,j} = \lambda^2 R(x_i, x_j)$ is the prior-precision matrix of the spline regression coefficients α_i solving the spline problem above (16).

Within this statistical formulation of the spline regression problem, the unknown parameter λ is *just* a prior-precision factor of the regression coefficients, which can be co-estimated by means of maximum likelihood from the data. The ML-estimation of the parameter λ also allows the spline to adapt to an increased information about the unknown function being approximated. That is, with more and more data, and therefore more and more information about the unknown function, the λ parameter will likely decrease and allow the spline to adapt to smaller features of the unknown function (if there are any). This flexibility of splines is a major advantage over parametric approaches like polynomials and piece-wise defined polynomials (e.g., B-splines), where the complexity of the function fit is determined by the a-priori chosen function class. That is, the degree of the polynomial or the number of knots in a B-spline.

Moreover, these regression coefficients α_i of the spline can be considered as random effects of an LMM. This allows to co-estimate regression splines together with *traditional* parametric fixed and random effects of LMMs within a single model, the so-called additive mixed models (AMMs, e.g., Wood, 2006).

Fitting splines to larger datasets, however, gets computationally expensive as the rank of the linear regression problem to be solved is equal to the sample size. Hence Wood developed a provable ideal, low-rank approximation approach to spline regression (Wood, 2003). This approach allows to keep the computational cost of a spline fit feasible as sample sizes grow. This approach requires an a-priori choice for the rank of the spline regression problem (the *k*-parameter). Although AMM implementations (e.g., mgcv, Wood, 2014) provide default values for the choice of *k*, one has to check whether a specific value *k* is sufficiently large.[7] In Wood (2006) a heuristic approach is presented: First a spline with some chosen *k* (e.g., the default) is fitted to the data. Then a second spline with a doubled rank 2*k* is fitted to the residuals of the first spline model. If the latter model finds a significant spline in the residuals of the first, the procedure is restarted with an increased value of *k* until no significant spline function can be found in the residuals.

References

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17–21. http://www.jstor.org/stable/2682899

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. http://CRAN.R-project.org/package=lme4 (R package version 1.1-6).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

---

[6]The precision matrix is the inverse of the covariance matrix and a precision function is a function that generates the precision matrix.

[7]Please note that *k* cannot be chosen too large as long as *k* is smaller than the sample-size. The computational complexity, however, grows cubic in *k*. To keep the computational costs low, one searches for the smallest *k* that is still sufficiently large to describe the data.

Berlinet, A., & Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Boston: Springer. https://doi.org/10.1007/978-1-4419-9096-9

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In Launer, R. L., & Wilkinson, G. N. (Eds.) *Army research office workshop on robustness in statistics* (pp. 201–236). New York: Academic Press.

Center for Human Resource Research. (1995). *National longitudinal study of youth: User's guide 1995*. Columbus, OH: Ohio State University.

Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, *19*(4), 915–922.

Demidenko, E. (2013). *Mixed models: Theory and applications*. Wiley-Interscience: Hoboken.

Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science*, *21*(3), 366–71. https://doi.org/10.1177/0956797610362060

Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, *59*(2), 127–136. https://doi.org/10.1198/000313005X41337

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, *2*(3), 235–247. https://doi.org/10.1037/1082-989X.2.3.235

Garcia-Marques, L., Garcia-Marques, T., & Brauer, M. (2014). Buy three but get only two: The smallest effect in a 2 × 2 ANOVA is always uninterpretable. *Psychonomic Bulletin & Review*, *21*(6), 1415–1430. https://doi.org/10.3758/s13423-014-0640-3

Hohenstein, S., Matuschek, H., & Kliegl, R. (2016). Linked linear mixed models: Simultaneous analyzing landing position and fixation duration in reading. *Psychonomic Bulletin & Review*, *24*(3), 637–651.

Kimeldorf, G. S., & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, *41*(2), 495–502.

Lewis, J. W., & Escobar, L. A. (1986). Suppression and enhancement in bivariate regression. *The Statistician*, *35*(1), 17–26. https://doi.org/10.2307/2988294

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, *107*(3), 385–393. https://doi.org/10.1037/0033-2909.107.3.385

Matuschek, H. (2016). *Applications of reproducing kernel Hilbert spaces and their approximations*. PhD Dissertation, University of Potsdam, Germany.

Matuschek, H., Kliegl, R., & Holschneider, M. (2015). Smoothing spline ANOVA decomposition of arbitrary splines: An application to eye movements in reading. *PLOS ONE*, *10*(3), e0119165. https://doi.org/10.1371/journal.pone.0119165

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. arXiv:1511.01864, https://doi.org/10.1016/j.jml.2017.01.001

McClelland, G., & Judd, C. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376–390. https://doi.org/10.1037/0033-2909.114.2.376

McClelland, G., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, type II errors, and false positive consumer psychology: Don't fight the power. *SSRN Electronic Journal*, *53*, 1–51. arXiv:1011.1669v3. https://doi.org/10.2139/ssrn.2580320

McConkie, G., Kerr, P., Reddix, M., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118. https://doi.org/10.1016/0042-6989(88)90137-X

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: http://www.r-project.org

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21–30.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, *47*(1), 1–52.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*(3), 524–536. https://doi.org/10.1037/0033-2909.109.3.524

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *45*(1), 133–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2006). *Generalized additive models*. Boca Raton: Chapman & Hall.

Wood, S. N. (2014). Mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. http://CRAN.R-project.org/package=lme4. (R package version 1.8-0).

Wood, S. N., Scheipl, F., & Faraway, J. J. (2012). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360. https://doi.org/10.1007/s11222-0129314-z

Wurm, L. H., & Fisicaro, S. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*(1), 37–48. https://doi.org/10.1016/j.jml.2013.12.003

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition*, *132*(2), 181–215. https://doi.org/10.1016/j.cognition.2014.03.008