Multilevel Modeling and the Explanatory Autonomy of Psychology

Abstract: This article argues for the explanatory autonomy of psychology drawing on cases from the multilevel modeling practice. This is done by considering a multilevel linear model in personality and social psychology, and discussing its philosophical implications for the reductionism debate in philosophy of psychology. I argue that this practice challenges the reductionist position in philosophy of psychology, and supports the explanatory autonomy of psychology.

1. Introduction

It has been a longstanding debate in philosophy concerning the status of psychology as an autonomous science (Jessor 1958; Putnam 1967, 1975; Fodor 1968, 1974, 1998; Block and Fodor 1972; Churchland 1989; Bickle 1998, 2003). While the debate was once focused on theory reduction that concerns reducing a higher-level theory's predicates to a lower-level theory's predicates via bridging laws or corresponding principles (Nagel 1961; Suppes 1962, 1967; Stegmüller 1976; Suppe 1977, 1989; Schaffner 1993), more recently it has shifted to a slightly different but closely related dispute: explanatory reductionism (Sarkar 1998; Bickle 2003; Weber 2005; Rosenberg 2006; Waters 2008).[1]

Explanatory reductionism is a position weaker than theory reductionism for all what it requires is only individual reductive explanation, conducted in a piecemeal manner, rather than wholesale theory reduction. More specifically, it is the view that a higher-level phenomenon can (always) be reductively explained by features residing at a lower-level (Kaiser 2015, 97).[2] Although different authors express the idea

---

[1] Note that when referring to levels I mean either hierarchical organization such as nations, provinces, cities, etc., or functional organization such as organisms, organs, tissues, etc.

[2] Sarkar (1998) distinguishes three different forms of reductionism—i.e., constitutive (ontological), epistemic and theory reductionism—and treats explanatory

differently, they share one core: to explain a given phenomenon of interest, an explanation that appeals exclusively to features at a lower-level is always better than an explanation that appeals to features at a higher-level. In philosophy of psychology, a lower-level may refer to the neural (molecular) level whereas a higher-level may denote the cellular, tissue, personal or social level, depending on what question is under consideration.[3]

This position is explicitly held by some key figures in the explanatory reductionist camp, e.g., John Bickle (2003), who manages to urge a 'ruthless' reduction of psychological phenomena (e.g., memory, consciousness and attention) to molecular neuroscience. For Bickle, the molecular level is clearly the desired level of analysis, a level from which phenomena happening at a higher-level—here the psychological level—can be reductively explained (other prominent proponents of this view are Clark 1980; Churchland 1989, 2007; Kim 1992).

---

reductionism as a special kind of epistemic reductionism. Maria Kaiser (2015) divides explanatory reductionism into two sub-types: (a) "a relation between a higher-level explanation and a lower-level explanation of the same phenomenon" (97); (b) individual reductive explanations, i.e., a relatively higher-level phenomenon can be reductively explained by a relatively lower-level feature (*Ibid*., 97). This essay will focus on the second sub-type.

[3] A parallel debate is happening in philosophy of biology (see, e.g., Weber 2005; Rosenberg 2006; Bechtel 2010).

Although Bickle's ruthless reduction has received serious criticisms (e.g., Craver 2005; Jong and Schouten 2005; Aizawa and Gillett 2009, 2011; Marshall 2009; Figdor 2010), one important line of argument, which is based on practice in psychology, has not been sufficiently explored in the literature.[4] This is the *multilevel linear modeling* practice (Mason et al. 1983; Goldstein 1995), also termed *mixed-effects modeling* (Singer 1998), *hierarchical linear modeling* (Raudenbush and Bryk 2002), etc. It has proven to be very useful and thus been widely employed in many disciplines of science, e.g., neuroscience (e.g., Gueorguieva and Krystal 2004; Baayen et al. 2008;

---

[4] Bourrat (2016) and Fang (2019) are two exceptions. Notice that this essay extends Fang (2019)'s arguments in many ways. First, Fang (2019) discusses how the multilevel modeling practice challenges explanatory reductionism in philosophy of science in general, but this essay examines how the practice casts doubt on explanatory reductionism in philosophy of psychology in particular. Second, this essay points out another two severe problems that the single-level analysis strategy faces, problems that are overlooked by Fang (2019) (see Section 2). Third, this essay further extends one argument made by Fang (2019) to philosophy of psychology as well as answers two possible objections, offers a new argument against explanatory reductionism, and deals with a new problem that any regression analysis must face (see Section 4). Finally, throughout this essay a number of examples drawn from personality and social psychology are discussed (see, especially, a case study in Section 3).

Aarts et al., 2014), biomedical research (e.g., Andersson-Roswall et al. 2010; Ard et al. 2015; Bilgel et al. 2016), behavioral and social sciences (Agresti et al. 2000; Berger and Tan 2004; Parzen et al. 2011), etc.

A multilevel linear model is one that simultaneously includes multiple (often nested or hierarchical) levels of analyses with explanatory variables ranging over several levels of organization to explain a phenomenon of interest. The appeal to multilevel rather than single-level models is grounded upon the fact that research, e.g., in the social sciences, always involves hierarchical data structures. For instance, a resident is typically nested within a district such that she is more similar to her neighbors in attitudes towards certain social phenomena, e.g., gender inequality, than to the residents from different districts in the same city.

I argue in this essay that the multilevel linear modeling practice constitutes a severe challenge to the explanatory reductionism position, for it flies in the face of the position's core. More specifically, it shows that appealing to the lower-level features cannot always provide a better explanation. Rather, scientists usually have to build multilevel models that incorporate both the lower-level and higher-level features to explain a phenomenon of interest. Moreover, the higher-level features sometimes even play a more important role in an explanation than their lower-level counterparts.

Before proceeding, let me say a few words on what I mean by 'explanatory autonomy'.[5] Explanatory autonomy refers to one particular type of autonomy that a

[5] I thank one anonymous referee for letting me clarify this notion.

science can possess. This type of autonomy is reflected by, and grounded upon, the fact that the explanations in a science are sufficient to explain the phenomena of interest in that field, and that there is no need to reduce the explanations provided by this science to some other explanations provided by some other, usually more fundamental sciences. Therefore, in this essay when I refer to 'the explanatory autonomy of psychology' I mean that the explanations in psychology are good enough to explain psychological phenomena such that there is no reason to reduce the psychological explanations to some other explanations.

The essay unfolds as follows. Section 2 examines the problems faced by the traditional single-level statistical methods. Section 3 briefly introduces the multilevel linear modeling practice happening in personality and social psychology. Section 4 discusses how the practice poses threat to the core of the explanatory reductionist position.

2. Problems with Traditional Statistical Methods

It goes without saying that one person's social behavior is not only influenced by her own personality but also by the social context in which she behaves. People are situated within social contexts, and different social contexts affect people differently. For example, people living in the same district in a metropolis tend to show more similar attitudes towards certain social phenomena, e.g., gender inequality, than people from different districts in the same metropolis. This common fact in social settings, namely hierarchical data structures,[6] violates one core assumption of traditional statistical methods: *independence of observation*, because knowing one person's attitude towards a social phenomenon in a district tells us something about the other people's attitudes in the same district. Traditional statistical methods employed in personality and social psychology, e.g., ordinary least squares, usually focus on a single-level analysis (e.g., either the individual or group-level) and overlook the nested/hierarchical data structure, which often results in mistaken analysis of the question under consideration (for a more comprehensive discussion of the shortcomings of the traditional statistical methods, see Nezlek 2001, 2008; Raudenbush and Bryk 2002; Christ et al. 2012; West et al. 2015).

As Fang (2019) has pointed out, the mistake of traditional statistical methods typically stems from two ways of accommodating hierarchical data structures: aggregation and disaggregation. On the one hand, one can take an aggregation strategy when encountering hierarchical data structures, namely, we base our analysis solely at the group-level. This can be done by, e.g., taking the means of individuals within a group as a group-level variable and then comparing the group means across different groups. In doing so, the subsequent analysis can be conducted at the group-

[6] A date set can be hierarchical in three different ways: nested/clustered data, repeated-measures data and longitudinal data. A nested data set is one that the units of analysis are nested within clusters of units. This essay concentrates on the nested data type; for other types see West et al. (2015, 9-15).

level. However, this move is problematic, because it leads one to interpret the relationships established between variables at the group-level as relationships between variables at the individual-level (Nezlek 2008, 843; also see Luke 2004). The problem here is that analysis of this sort is unable to properly capture possible within-group variability, and thus unable to make precise prediction about each individual.

On the other hand, one can take an alternative strategy, disaggregation. Namely, we base our analysis solely at the individual-level. This can be done, e.g., by pooling all group-level variability into a single error term in the model. This move is also problematic, however, because individuals within the same group tend to have correlated errors, which violates the independence assumption. Worse still, this move usually leads one to reduce all the group-level characteristics to the individual-level and thereby misleadingly interprets relationships established at the individual-level as relationships at the group-level (Nezlek 2008, 843). In other words, features such as group size idiosyncratic to each group and affecting the relationships between variables at the individual-level are simply ignored.

Social psychologists Wagner et al. (2006)'s study of the relationship between the percentage of minorities in an area and majority members' attitudes towards them illustrates the points made above. They found that with the same data set at hand, shifting the level of analysis from one to another would result in contradictory results. For example, if they focused on a lower-level of analysis, e.g., different districts within a nation, it would turn out that a higher percentage of minorities is associated with more positive intergroup attitudes. By contrast, if they shifted their focus to a higher-level of analysis, e.g., different nations, it would turn out that a higher percentage of minorities is associated with more negative intergroup attitudes.

However, in addition to violating the independence of observation assumption and mistakenly interpreting relationships established at one level as relationships at another, there are other pressing issues for a single-level analysis that Fang (2019) does not discuss. The first one is related with random errors, as briefly mentioned above. In traditional single-level analysis such as ordinary least squares regression, random errors are assumed to be independent (and normally distributed). When the data set involves hierarchical structures, however, the independence assumption of residuals is also violated. This is because "the individual-level residuals are dependent within each unit because they are common to every individual within that unit" (Heck and Thomas 2015, 30). This has implications for other aspects of modeling building, e.g., intercepts and slopes, as will be shown in the next section.

The second is related with underestimated standard errors. Recall that people living in the same district tend to be more similar to one another in important features, e.g., social attitudes, than people from different districts. Technically put, this entails that "the mean correlation of variables measured for persons of one group is higher than the average correlation between variables measured in individuals from different groups" (Christ et al. 2012, 244). In other words, the nested data structure usually generates correlated errors. Therefore, a single-level analysis, due to the failure of taking into account these correlated errors, usually obtains estimates of standard errors that are smaller than they should be. Moreover, the effect of obtaining underestimated standard errors carries over to hypothesis test, for the latter relies on the former (Heck and Thomas 2015, 30).[7]

Given these challenging problems facing the single-level analysis framework, one option is to find ways to fix the framework. However, many such attempts have turned out to be unsatisfactory since they usually engender other more recalcitrant problems (Luke 2004, 7; see also Nezlek 2008; Heck and Thomas 2015). A different strategy, as has been taken by mainstream psychologists, is to put the single-level analysis strategy aside and adopt the new emerging multilevel modeling framework. The next section will briefly describe this new strategy using cases from psychology.

3. Multilevel Linear Modeling in Psychology: A Brief Introduction

Thanks to the innovations in statistical analysis during the last three decades, psychologists now are able to deal with situations with hierarchical data structures using multilevel linear models. A multilevel linear model, also called a mixed-effects model, has a mix of random and fixed effects (these terms will be explained below). The difference between a mixed-effects model and a fixed-effects model relies on the fact that the former attempts to capture random or stochastic variability in the data arising from different sources, e.g., people and/or the districts they live in, while the

---

[7] More specifically, this is because the hypothesis test relies on the ratio of a parameter to its standard error.

latter does not do so. For example, in mixed-effects modeling, e.g., different social contexts can have different effects on how one person's personality impacts her social behavior. As a result, the effects of social context are treated as random across different contexts in a mixed-effects model, whereas in a fixed-effects model the effects of social context are treated as fixed, namely, people's personality impacts their social behaviors in the same way across various distinct contexts.

To capture this kind of random or stochastic variability in the data, a multilevel linear model *must* simultaneously incorporate multiple levels of analyses. To show this, consider an example drawn from personality and social psychology research. Kuppens et al. (2008) attempted to explore the relationship between the frequency of positive and negative emotions (FE) people experienced and their level of life satisfaction (LLS) in different nations. They hypothesized that different national cultures (NC), e.g., survival/self-expression and individualism/collectivism, might moderate the relationship differently. More specifically, they hypothesized that "individualism would moderate primarily the relation between life satisfaction and negative emotions and that survival/self-expression would primarily moderate the relation between positive emotions and life satisfaction" (2008, 72). To test these hypotheses, the first step is to build a multilevel linear model. This model has two levels of analysis: level-1 for people ($i = 1, 2, \ldots, N$) who are clustered in level-2 nations ($j = 1,2, \ldots, n$). The relationship between the predictor variable FE and the dependent variable LLS is described below:

$$LLS_{ij} = \beta_{0j} + \beta_{1j}FE_{ij} + \varepsilon_{ij} \quad (1)$$

In equation (1), LLS_{ij} stands for the level of life satisfaction for person i in the jth nation; β_{0j} denotes a level-1 intercept that indicates the mean of life satisfaction for the jth nation; β_{1j} refers to a level-1 slope for the predictor FE_{ij} that represents different effects of emotions on life satisfaction; and the residual ε_{ij}, an error term, represents the deviation of person i's level of life satisfaction from the level-2 mean in the jth nation. ε_{ij} is usually assumed to be normally distributed with a mean of zero and a constant level-1 variance, σ^2. At the first glance, equation (1) strikes us as a multiple regression model. Looking closely, however, the subscript j unmasks the fact that there is a group-level component included in the model.

Unlike traditional statistical methods where intercept and slope are routinely treated as fixed for people across different nations, here in equation (1) the intercept β_{0j} and slope β_{1j} are treated as random, namely, they can vary across the level-2 units (i.e., nations) such that different nations can have their own distinct intercepts and slopes. In technical terms, this means that we can treat β_{0j} and β_{1j} at level-1 as dependent variables of level-2 predictors. This idea is expressed in the following two equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}NC_j + u_{0j} \quad (2)$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11} NC_j + u_{1j} \tag{3}$$

In equations (2), β_{0j} denotes the level-1 intercept in level-2 unit j, i.e., unit j's own intercept; γ_{00} stands for the average intercept across the level-2 units, controlling for the level-2 predictor, NC_j (i.e., national cultures); γ_{01} the effect (slope) of the level-2 predictor, NC_j; and u_{0j} is the error for unit j. In equations (2), β_{1j} denotes the level-1 slope in level-2 unit j, i.e., unit j's own slope; γ_{10} the average slope across the level-2 units, controlling for the level-2 predictor, NC_j; γ_{11} is the effect of the level-2 predictor, NC_j; and u_{1j} the error for unit j.

We have seen in equations (2) and (3) "how each of the level-1 parameters are functions of level-2 predictors and variability" (Luke 2004, 10). By way of including one level-2 predictor into the model (plus its associated error terms), these equations are able to capture the variations in the intercepts and slopes that traditional single-level analysis frameworks cannot capture. This treatment is grounded upon the recognition that higher-level features such as national cultures may mediate the lower-level relationship between, e.g., FE and LLS. In other words, "the magnitude of a relationship observed within groups is dependent on contextual or organizational features defined by higher-level units" (Heck and Thomas 2015, 43). The higher-level's influence on the lower-level relationship is termed a *cross-level interaction*, defined as "the potential effects variables at one level of a data hierarchy have on relationships at another level" (Ibid., 42).

Substituting the level-2 components, i.e., equations (2) and (3), of the model into the level-1 equation (1) yields the following combined equation:

$$LLS_{ij} = \left[\gamma_{00} + \gamma_{10}FE_{ij} + \gamma_{01}NC_j + \gamma_{11}FE_{ij}NC_j\right] + \left[u_{1j}FE_{ij} + u_{0j} + \varepsilon_{ij}\right] \qquad (4)$$

Equation (4) looks a bit complex, but it is simply composed of two constituents: the fixed-effects constituent included in the left square bracket, and the random-effects constituent included in the right. It now becomes clear why it is called a mixed-effects model—simply because the model is a mix of a fixed-effects part and a random-effects part. One term in the equation deserves special attention: $\gamma_{11}FE_{ij}NC_j$, referring to a cross-level interaction between level-1 and level-2 variables, as defined above.

This model has a couple of parameters to estimate (four fixed effects in the left square bracket and three random effects in the right). To test this model, Kuppens et al. (2008) examined a data set on the relationship between FE and LLS provided by the International College Survey 2001. This data set features a hierarchical/nested nature for it involved 8,557 participants coming from 46 different nations. Using multilevel linear modeling techniques as described above, Kuppens and colleagues (2008) were able to conclude that the modeling results unambiguously confirmed their expectations:

"Only individualism moderated the relation between negative emotions and life

> satisfaction: The more a national culture stresses individualistic values, the more adverse is the impact of negative emotions on life satisfaction. With respect to positive emotions, only survival/self-expression moderated their relation to satisfaction with life: The more a nation stresses self-expression values, the stronger the impact of positive emotions on the life satisfaction of its inhabitants." (2008, 72)

This example has illustrated the general point that situations involving hierarchical data structures routinely call for hierarchical modeling treatments. Without such a kind of treatment, variations across groups might simply go unnoticed or dismissed, leading to those remarkable problems pointed out in Section 2. This hierarchical modeling practice has important implications for the philosophical debate over reductionism versus anti-reductionism, for it challenges the core championed by many reductionists in the philosophy of psychology.

4. Discussion: Why Explanatory Reductionism Is Questionable

Fang (2019) has convincingly demonstrated that an explanation sometimes must appeal to both the lower-level and higher-level features in order to satisfactorily explain a phenomenon of interest; the lower-level alone is not always sufficient to explain. However, though Fang (2019)'s idea is very general such that it can be applied to many branches of science where the multilevel modeling practice is taking place, he does not show how that practice also poses a threat to the reductionist position in philosophy of psychology in particular. So, to fill in that gap, I will further argue in what follows that, first, to explain a psychological phenomenon of interest we sometimes also must appeal to both the lower-level and the higher-level features, and second, more seriously, the higher-level features of an explanation sometimes can contribute even more to explaining a psychological phenomenon of interest than their lower-level counterparts. Note that the first argument is an extension of Fang (2019)'s, but the second is brand-new and offers further support to Fang (2019). Finally, I will deal with one potential problem that any regression analysis—including the multilevel modeling framework—must face: the *multi-collinearity* problem.

4.1. Multiple levels are required

Philosophers such as John Bickle (2003) believe that psychological phenomena like attention can be simply explained at the neural level. What they assume is that a single-level (usually a lower-level) analysis is sufficient to explain a psychological phenomenon of interest. However, as pointed out in Section 2, while it might be true that the occurrence of a psychological phenomenon has its neural underpinnings, a single-level strategy invoking only neural factors may fail to explain the phenomenon when the phenomenon involves a hierarchical data structure; rather, to explain a phenomenon of interest, multiple levels are required. Forcefully reducing all relevant levels of analyses to a single-level (e.g., neural) may incur the various serious problems discussed in Section 2.

To look closely at how reducing all relevant levels of analyses to a single-level might induce severe problems, and to see how multiple levels are required to explain a phenomenon of interest when a hierarchical data structure is present, consider an example from personality and social psychology. Many people used to believe that children's innate differences in ability to acquire language accounted for their differences in language development. Naturally, this supposition encouraged a single-level analysis strategy where children's innate differences in ability to acquire language, as represented by their parents' scores on standardized vocabulary tests, were treated as the predictor variable while children's differences in language development as the dependent variable. However, heritability studies showed that children's innate differences can only explain 10-20% of the variance in children's scores on the same tests as their parents did (Raudenbush and Bryk 2002, 8). This means that there must be some missing factors that could potentially account for the rest of the variance in the children's scores. To capture the missing variance, an alternative multilevel linear model was suggested (Huttenlocher et al. 1991), which incorporated another level of analysis that is related to children's exposure to language during infancy. Interestingly, it turned out that exposure to language played a much more significant role in children's language development than initially thought, which can account for most of the variance in the children's scores.

This example, plus the one discussed in the last section, have clearly demonstrated that multiple levels are required to account for a phenomenon of interest when the phenomenon involves a hierarchical data structure. Some might point out that the higher-level factors in each example do account for a large proportion of the variance, and therefore a single-level analysis strategy invoking only the higher-level might be not so bad at explaining at all. However, remember how we know the fact that the higher-level factors account for most of the variance; we know this only when we have already built a multilevel model that properly decomposes the variance into different components. Without building such a multilevel model, it even makes no sense to talk about the proportion of variance. Moreover, it is important to note that what a reductionist typically aims at is not an explanation invoking only the higher-level factors, but rather an explanation resorting exclusively to the lower-level factors (we will return to this point in the next section).

Before going to the next section, let us consider another possible objection. One might point out that the children's language development example does not directly challenge Bickle's proposal, for the example does not invoke anything like neurons. Bickle's proposal concerns whether a psychological phenomenon can be explained solely at the neural level whereas our example concerns whether children's language development can be fully explained by the children's innate ability. They are talking past each other. Hence, as the argument goes, it remains open whether we can find phenomena that can be exclusively explained at the neural level.[8] Nevertheless, although the language development example may only hint at the possible limitations

---

[8] I thank Paul Roth for letting me notice this potential line of objection.

of Bickle's proposal, there is evidence showing directly that neuroscientific research is also vulnerable to the nested data structure problem so that multilevel models are better preferred when the date set is nested (Lazic 2010; Aarts et al. 2014; Aarts et al. 2015). Aarts and colleagues, after surveying 314 research articles published in prestigious journals *Science*, *Nature*, *Cell*, *Nature Neuroscience* and *Neuron* in 2012 and the first six months of 2013, obtain the result that at least 53% of these articles included nested data (Aarts et al. 2014).

To illustrate how the nested nature of data affects neuroscientists' choice of analytic strategies, consider an example discussed by neuroscientists Aarts et al. (2015). We might be interested in investigating the differences in the number of vesicles in presynaptic boutons between wild type (WT) and knock-out (KO) mice, since, suppose, the differences may explain some observed differences in mice's behaviors. Because each neuron has a number of presynaptic boutons, measuring the number of vesicles in the set of boutons of each neuron leads to multiple measurements within each neuron. All the measurements from the same neuron are assumed to be obtained from the same environmental condition (e.g., the same genotype), though the environmental conditions may vary from neuron to neuron. This means that the number of vesicles in the set of boutons within the same neuron may be more similar than that from different neurons (e.g., boutons within the same neuron may have the same number of vesicles while boutons from different neurons may have different numbers of vesicles). Hence, the measurements/observations within each neuron are not independent. In other words, the measurements are nested within neurons, resulting in the hierarchical data structure in which the number of vesicles within each bouton constitutes the level-1 variable while the neuron constitutes the level-2 variable. The scenario can become even more complicated when the neurons themselves are situated within varying subcellular, cellular or tissue environmental conditions (of the same animal), where a still higher-level clustering variable (i.e., the level-3) representing these varying environmental conditions might be introduced. Examples of this latter situation can be easily found in clinical and preclinical neuroscience (Raudenbush et al. 2000; Lazic and Essioux 2013).

As Aarts et al. (2015) claim, settings like these are very common in neuroscience, and any single-level strategy that overlooks this nested data structure may fall prey to committing all those problems discussed in Section 2. Therefore, the moral obtained from these examples is that, *contra* Bickle, explaining higher-level (e.g., psychological) phenomena using resources solely from the molecular (or neuronal) level is not always feasible due to the common nested data structure. Rather, we usually have to appeal to both the lower-level and higher-level features to explain a phenomenon of interest. Notice that in the children's language development example discussed above, the higher-level features explain a larger proportion of the variance in the children's scores—this leads us to the next argument below.

4.2. Higher-level features contribute more

For a reductionist, higher-level features are something to be reduced rather than something to be sought after. Correspondingly, for them, lower-level features are *always* better than their higher-level rivals with respect to explaining a phenomenon of interest. However, our case study in Section 3 has shown that all levels of analyses are equally important in dealing with situations with hierarchical data structures. There is no point to claim which level is more important. Nevertheless, there might be a sense in which one can reasonably claim, contrary to a reductionist' expectation, that higher-level features *sometimes* do play a much more important role in explaining a phenomenon of interest. This, again, relates to the decomposition of variance.

To see this, consider a research conducted by Bryk and Raudenbush (1988). They collected a longitudinal data set from 618 students distributed over 86 schools,[9] and repeatedly measured each student's mathematics achievement on five different occasions. They thus built a three-level linear model to explain children's mathematics achievement at school, with the first being the occasion-level, the second being the student-level and the third the school-level. This multilevel analysis strategy had the advantage of decomposing the variations in student's mathematics achievement into within-school and between-school components. With the aid of this strategy, they found that 83% of the variance in student's mathematics achievement was due to between-school effects. In other words, the variance was largely stemmed from the higher-level rather than the lower-level, and thus the variations in student's

[9] A longitudinal data set is obtained in such a way that "subjects are measured repeatedly over time or under different occasions" (West et al. 2015, 1).

mathematics achievement were largely accounted for by the higher-level features.

One might object that this example merely shows that higher-level features *sometimes* can contribute more to accounting for the variance in a target phenomenon. So, this leaves room for the possibility that *at other times* lower-level features might contribute more. Nonetheless, even though this might be true, it does not weaken my argument. First, remember how we can determine which level of features is more important in the first place. As pointed out in Section 4.1, we know which level of features contributes more to accounting for the variance only when we have already built a multilevel model. So, no matter how important the lower-level features turn out to be, in order to know which level is explanatorily more important a multilevel model must be in place from the very beginning. Second, notice that a reductionist position is not as weak as something that lower-level features are *sometimes* more important, but rather that lower-level features are *always* more important. The reductionist aims high, to be sure. That is why philosophers such as John Bickle aim for a non-negotiable 'ruthless' reduction of psychological phenomena to molecular neuroscience. Yet, if what we have said about the multilevel modeling practice is right, then it has at least demonstrated that the reductionist's dream may turn sour, for higher-level features sometimes can contribute even more to explaining a phenomenon of interest, even though lower-level features also play a role in that explanation. Therefore, unless the reductionist is ready to give in and opt for a much weaker position, the example still constitutes an effective objection.

Thus, we have shown that, against the reductionist's expectation, higher-level features sometimes can even contribute more to explaining a phenomenon of interest. This requires us to take a *pluralist* stance—rather than a somewhat *monist* viewpoint favoring a particular level of analysis—towards scientific explanation, a stance that is at least able to accommodate the fact that higher-level features can sometimes even more important than their lower-level counterparts in an explanation.

4.3. The multi-collinearity problem[10]

The multi-collinearity problem occurs when many (i.e., often more than two) independent variables in a regression model are correlated. This is problematic because, by designing a regression analysis, we only want to establish the dependence relationship between the independent variables and the dependent variable, and therefore to isolate each independent variable's influence on the dependent variable. When this problem arises, changes in one independent variable in the model will always be correlated with changes in the other independent variables, thus rendering it impossible or at least very difficult to independently estimate the relationship between each independent variable and the dependent variable. Hence, the detrimental impact of multi-collinearity carries over to coefficient estimates, their associated standard errors, as well as variance components.[11]

---

[10] I thank one anonymous referee for helping me notice this problem.

[11] For a detailed discussion of these effects on a hierarchical linear model, see Shieh

Though this is definitely a serious problem, it does not follow that it can undermine what I have said about the multilevel modeling practice, nor can it weaken the philosophical implications for the reductionist position in philosophy of psychology. First of all, the problem of multi-collinearity is so general and prevalent that any regression analysis must face it, be it a single-level analysis or a multilevel analysis (Bingham and Fry, 2010, 174-177; also see Greene, 2012). So, it does not put any less threat to the traditional single-level analysis methods than to the multilevel modeling practice. If it is a problem, it is a problem for all of us.

Second, the good news is that scientists have found ways to detect and at least partially resolve the problem. One very effective way to detect multi-collinearity, for example, is via the *variance inflation factor* (VIF), which can help pinpoint the correlation between independent variables as well as measure the strength of the correlation (for details of this method, see James et al. (2013, 101-102)). However, though it is relatively easy to detect multi-collinearity, to resolve it is usually not so straightforward. The most popular method to cope with this problem is called *ridge regression*, in which "the matrix of covariances among the independent variables is modified slightly in order to reduce the standard errors of the partial regression coefficients" (Allen 1997, 178). There are also other methods suggested in the literature, for instance, we re-collect data in such a way that the problem simply disappears, or re-define our research problem if we find it reasonable to do so, or re-

---

and Fouladi (2003).

specify the multiple regression model by deleting an independent variable from the model, etc. (Yu et al. 2015, 123). However, we must note that all these methods may only suit certain rather than all research purposes and that there is no one-size-fits-all method.[12]

Therefore, multi-collinearity does not seem to cause an irremediable problem for the multilevel modeling practice, nor does it count as a challenge only to the multilevel modeling practice.

5. Conclusion

This article has demonstrated that the multilevel explanatory strategy, in coping with complex social settings with hierarchical data structures, constitutes serious challenges to the explanatory reductionist position advocated by many philosophers. It is so challenging because it flies in the face of the core of explanatory reductionism: to explain a given phenomenon of interest, an explanation that appeals exclusively to

[12] Also note that scientists usually do not worry about the existence of multi-collinearity but rather the degree or strength of multi-collinearity. Therefore, when the strength of multi-collinearity is relatively low (e.g., below 5 in terms of the variance inflation factor), when the research purpose is prediction rather than explanation, or when the correlated variables are not what scientists are interested in, multi-collinearity only constitutes a benign problem that can be simply ignored.

features at a lower-level is always better than an explanation that appeals to features at a higher-level. As a consequence, the multilevel modeling practice is in support of the explanatory autonomy of psychology, for it shows that higher-level features considered by psychology, e.g., personality and social features, are not necessarily reduced when explaining a psychological phenomenon of interest. In other words, the explanations in psychology are good enough to explain psychological phenomena such that there is no reason to reduce the psychological explanations to some other explanations; reducing such multilevel explanations can only cripple them rather than make them better.

References

Aarts, Emmeke, Matthijs Verhage, Jesse V. Veenvliet, Conor V. Dolan, and Sophie Van Der Sluis. 2014. "A Solution to Dependency: Using Multilevel Analysis to Accommodate Nested Data." *Nature Neuroscience* 17 (4): 491.

Aarts, Emmeke, Conor V. Dolan, Matthijs Verhage, and Sophie van der Sluis. 2015. "Multilevel Analysis Quantifies Variation in the Experimental Effect While Optimizing Power and Preventing False Positives." *BMC Neuroscience* 16 (1): 94.

Agresti, Alan, James G. Booth, James P. Hobert, and Brian Caffo. 2000. "Random-Effects Modeling of Categorical Response Data." *Sociological Methodology* 30 (1): 27–80.

Aizawa, Kenneth, and Carl Gillett. 2009. "The (Multiple) Realization of Psychological and Other Properties in the Sciences." *Mind & Language* 24 (2): 181–208.

Aizawa, Ken, and Carl Gillett. 2011. "The Autonomy of Psychology in the Age of Neuroscience." In *Causality in the Sciences*, by Phyllis McKay Illari, Federica Russo, and Jon Williamson, 6:202–223. Oxford: Oxford University Press.

Allen, Michael Patrick. 1997. *Understanding Regression Analysis*. Boston, MA: Springer.

Andersson-Roswall, Lena, Elisabeth Engman, Hans Samuelsson, and Kristina Malmgren. 2010. "Cognitive Outcome 10 Years after Temporal Lobe Epilepsy Surgery: A Prospective Controlled Study." *Neurology* 74 (24): 1977–1985.

Ard, M. Colin, Nandini Raghavan, and Steven D. Edland. 2015. "Optimal Composite Scores for Longitudinal Clinical Trials under the Linear Mixed Effects Model." *Pharmaceutical Statistics* 14 (5): 418–426.

Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. 2008. "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59 (4): 390–412.

Bechtel, William. 2010. "The Downs and Ups of Mechanistic Research: Circadian Rhythm Research as an Exemplar." *Erkenntnis* 73 (3): 313–28.

Berger, Martijn PF, and Frans ES Tan. 2004. "Robust Designs for Linear Mixed Effects Models." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53 (4): 569–581.

Bickle, John. 1998. *Psychoneural Reduction: The New Wave*. Cambridge: MIT Press.

Bickle, John. 2003. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Springer.

Bilgel, Murat, Jerry L. Prince, Dean F. Wong, Susan M. Resnick, and Bruno M. Jedynak. 2016. "A Multivariate Nonlinear Mixed Effects Model for Longitudinal Image Analysis: Application to Amyloid Imaging." *Neuroimage* 134: 658–670.

Bingham, Nicholas H., and John M. Fry. 2010. *Regression: Linear Models in Statistics*. Springer Science & Business Media.

Block, Ned J., and Jerry A. Fodor. 1972. "What Psychological States Are Not." *The Philosophical Review* 81 (2): 159–181.

Bourrat, Pierrick. 2016. "Generalizing Contextual Analysis." *Acta Biotheoretica* 64

(2): 197–217.

Bryk, Anthony S., and Stephen W. Raudenbush. 1988. “Heterogeneity of Variance in Experimental Studies: A Challenge to Conventional Interpretations.” *Psychological Bulletin* 104 (3): 396–404.

Christ, Oliver, Chris G. Sibley, and Ulrich Wagner. 2012. “Multilevel Modeling in Social and Personality Psychology.” In *The Oxford Handbook of Personality and Social Psychology*, by Kay Deaux and Mark Snyder, 240–261. Oxford: Oxford University Press.

Churchland, Paul M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: MIT press.

Churchland, Paul. 2007. *Neurophilosophy at Work*. Cambridge: Cambridge University Press.

Clark, Austen. 1980. *Psychological Models and Neural Mechanisms: An Examination of Reductionism in Psychology*. Oxford: Clarendon Press.

Craver, Carl F. 2005. “Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience.” *Studies in History and Philosophy of Science Part C* 36 (2).

Fang, Wei. 2019. “Mixed-Effects Modeling and Non-Reductive Explanation.” *Philosophy of Science*, forthcoming.

Figdor, Carrie. 2010. “Neuroscience and the Multiple Realization of Cognitive Functions.” *Philosophy of Science* 77 (3): 419–456.

Fodor, Jerry A. 1968. *Psychological Explanation: An Introduction to the Philosophy *of Psychology*. New York: Crown Publishing Group/Random House.

Fodor, Jerry A. 1974. “Special Sciences (or: The Disunity of Science as a Working Hypothesis).” *Synthese* 28 (2): 97–115.

Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Goldstein, Harvey. 1995. *Multilevel Statistical Models*. New York: John Wiley & Sons.

Greene, William H. 2012. *Econometric Analysis, 7th*. Pearson.

Gueorguieva, Ralitza, and John H. Krystal. 2004. “Move over Anova: Progress in Analyzing Repeated-Measures Data And its Reflection in Papers Published in the Archives of General Psychiatry.” *Archives of General Psychiatry* 61 (3): 310–317.

Heck, Ronald H., and Scott L. Thomas. 2015. *An Introduction to Multilevel Modeling Techniques*. 3rd ed. Routledge.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons. 1991. “Early Vocabulary Growth: Relation to Language Input and Gender.” *Developmental Psychology* 27 (2): 236.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. New York: Springer Science & Business Media.

Jessor, Richard. 1958. “The Problem of Reductionism in Psychology.” *Psychological Review* 65 (3): 170.

Jong, Looren de Huib, and Maurice KD Schouten. 2005. “Ruthless Reductionism: A

Review Essay of John Bickle's Philosophy and Neuroscience: A Ruthlessly Reductive Account." *Philosophical Psychology* 18 (4): 473–486.

Kaiser, Marie I. 2015. *Reductive Explanation in the Biological Sciences*. Springer.

Kim, Jaegwon. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52 (1): 1–26.

Kuppens, Peter, Anu Realo, and Ed Diener. 2008. "The Role of Positive and Negative Emotions in Life Satisfaction Judgment across Nations." *Journal of Personality and Social Psychology* 95 (1): 66–75.

Lazic, Stanley E. 2010. "The Problem of Pseudoreplication in Neuroscientific Studies: Is It Affecting Your Analysis?" *BMC Neuroscience* 11 (1): 5.

Lazic, Stanley E., and Laurent Essioux. 2013. "Improving Basic and Translational Science by Accounting for Litter-to-Litter Variation in Animal Models." *BMC Neuroscience* 14 (1): 37.

Luke, Douglas A. 2004. *Multilevel Modeling*. Vol. 143. London: SAGE Publications.

Marshall, Peter J. 2009. "Relating Psychology and Neuroscience: Taking up the Challenges." *Perspectives on Psychological Science* 4 (2): 113–125.

Mason, William M., George Y. Wong, and Barbara Entwisle. 1983. "Contextual Analysis through the Multilevel Linear Model." *Sociological Methodology* 14: 72–103.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt: Brace & World.

Nezlek, John B. 2001. "Multilevel Random Coefficient Analyses of Event- and

Interval-Contingent Data in Social and Personality Psychology Research." *Personality and Social Psychology Bulletin* 27 (7): 771–85.

Nezlek, John B. 2008. "An Introduction to Multilevel Modeling for Social and Personality Psychology." *Social and Personality Psychology Compass* 2 (2): 842–860.

Parzen, Michael, Souparno Ghosh, Stuart Lipsitz, Debajyoti Sinha, Garrett M. Fitzmaurice, Bani K. Mallick, and Joseph G. Ibrahim. 2011. "A Generalized Linear Mixed Model for Longitudinal Binary Data with a Marginal Logit Link Function." *The Annals of Applied Statistics* 5 (1): 449.

Putnam, Hilary. 1967. "Psychological Predicates." In *Art, Mind, and Religion*, edited by William H. Capitan and Daniel D. Merrill, 1:37–48. University of Pittsburgh Press.

Putnam, Hilary. 1975. "Philosophy and Our Mental Life." In *Mind, Language, and Reality*. Cambridge University Press.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. London: SAGE Publications.

Raudenbush, Stephen W., and Xiaofeng Liu. 2000. "Statistical Power and Optimal Design for Multisite Randomized Trials." *Psychological Methods* 5 (2): 199.

Rosenberg, Alexander. 2006. *Darwinian Reductionism: Or, How to Stop Worrying and Love Molecular Biology*. University of Chicago Press.

Schaffner, Kenneth F. 1993. *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Shieh, Yann-Yann, and Rachel T. Fouladi. 2003. “The Effect of Multicollinearity on Multilevel Modeling Parameter Estimates and Standard Errors.” *Educational and Psychological Measurement* 63 (6): 951–985.

Singer, Judith D. 1998. “Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models.” *Journal of Educational and Behavioral Statistics* 23 (4): 323–355.

Stegmüller, Wolfgang. 1976. *The Structure and Dynamics of Theories*. New York: Springer.

Suppe, Frederick. 1977. *The Structure of Scientific Theories*. Urbana, Illinois: University of Illinois Press.

Suppe, Frederick. 1989. *The Semantic Conception of Theories and Scientific Realism*. Urbana, Illinois: University of Illinois Press.

Suppes, Patrick. 1962. “Models of Data.” In *Logic, Methodology, and the Philosophy of Science*, by Ernest Nagel, Patrick Suppes, and Alfred Tarski, 24–35. Stanford University Press.

Suppes, Patrick. 1967. “What Is a Scientific Theory?” In *Philosophy of Science Today*, by Sidney Morgenbesser. New York: Meridian Books.

Wagner, Ulrich, Oliver Christ, Thomas F. Pettigrew, Jost Stellmacher, and Carina Wolf. 2006. “Prejudice and Minority Proportion: Contact Instead of Threat Effects.” *Social Psychology Quarterly* 69 (4): 380–390.

Waters, C. Kenneth. 2008. “Beyond Theoretical Reduction and Layer-Cake Antireduction: How DNA Retooled Genetics and Transformed Biological

Practice.” In *The Oxford Handbook of Philosophy of Biology*, by Michael Ruse, 238–262. Oxford: Oxford University Press.

Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

West, Brady T., Kathleen B. Welch, and Andrzej T. Galecki. 2015. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: CRC Press.

Yu, Han, Shanhe Jiang, and Kenneth C. Land. 2015. “Multicollinearity in Hierarchical Linear Models.” *Social Science Research* 53: 118–136.