Political Behavior, Vol. 24, No. 2, June 2002 (© 2002)

PARENTAL SOCIALIZATION AND RATIONAL PARTY IDENTIFICATION

Christopher H. Achen

This article constructs a rational choice model of the intergenerational transmission of party identification. At a given time, identification with a party is the estimate of average future benefits from candidates of that party. Experienced voters constantly update this expectation using political events since the last realignment to predict the future in accordance with Bayes Rule. New voters, however, have no experience of their own. In Bayesian terms, they need prior beliefs. It turns out that under certain specified conditions, these young voters should rationally choose to employ parental experience to help orient themselves to politics. The resulting model predicts several well–known features of political socialization, including the strong correlation between parents' and children's partisanship, the greater partisan independence of young voters, and the tendency of partisan alignments to decay.

Key words: socialization; party identification; political parties; party systems; Bayesian; retrospective voting.

> All I know is we're not Republicans. My father isn't.—Judith, age 10 (Greenstein, 1969, p. 23)

INTRODUCTION

The predictive power of "party identification" in American elections is well-nigh overwhelming, and, after some initial setbacks, evidence has accumulated that, properly measured, it does well elsewhere, too (Converse and Pierce, 1985, 1986; Green, Palmquist, and Schickler, 2002; Johnston, 1988; Miller and Shanks, 1996; Shively, 1980). Partisanship remains the central factor in explaining not just how people vote, but also how they see the political world, just as Campbell, Converse, Miller, and Stokes (1960/1980) asserted 40 years ago and as sophisticated recent investigations have confirmed (Bartels, 2001b).

Christopher H. Achen, Institute for Social Research—Center for Political Studies, University of Michigan, 426 Thompson St., P.O. Box 1248, Ann Arbor, MI 48106-1248 (achen@umich.edu).

The structuring power of partisanship in elections makes the study of partisan socialization important. As many scholars have noted, if every generation began anew, democratic electorates would be less stable and more susceptible to antisystem demagogues (for example, Easton and Dennis, 1969, chaps. 2, 3, 19). Longstanding party attachments inoculate citizens against overnight new party movements, some of which may harbor potential totalitarians. Converse (1964) pointed to weak partisanship, lacking the strength that builds up over generations, as a contributing factor in Hitler's rise to power via the ballot box. Even historians with no ties to the "Michigan model" of party identification have made a similar argument, noting particularly the weak party ties of the German right in the electorate, many of whose voters Hitler converted to his cause (Kershaw, 1983, pp. 25–26). Although the study of the transmission of partisanship between generations has engaged only a few political scientists, its intellectual importance is considerable. So far as contemporary scholarship can determine, if Hitler matters, then so does the study of partisan socialization.

Happily, partisanship in the electorate is well established for much of the citizenry in most successful democracies, since most parents transmit their party identification to their children. Though expressed partisanship is not always meaningful among the very young (Greenstein, 1965, chap. 4; Hess and Torney, 1968, pp. 103–104), careful longitudinal study of Americans has shown that by adolescence, most children have a partisan identification connected to political preferences in the same manner as adults, though not always as strongly. Among those children who have a partisan preference, nearly all share it with their parents (Jennings and Niemi, 1981, pp. 89–91). Similar findings generally hold in stable European democracies (see Sears, 1975 for a review). These results are so familiar and well established that they are often used as the explanatory axioms for topics as diverse as the growth of partisanship in new electorates (Converse, 1969) and the origins of the American gender gap in differential childhood socialization (Trevor, 1999).

Yet the very familiarity of parent–child partisan agreement has often let parental success go unquestioned. The social-psychological tradition has too often treated the child as the passive recipient of parental choices. But what child was ever like that? Parents are rarely able to influence their teenage children's hairdos, clothing styles, tastes in popular music, or even more important decisions such as the choice of a life partner. Neither the trivial nor the consequential seem to be under parental control. Why should party identification be any different? Put more precisely, why do teenagers implicitly accept their parents' advice about political parties while they avoid taking it on a great many other topics?

To answer these questions, and to help unify the main empirical findings in the literature, this article constructs a rational choice model of political socialization. Here the child is no longer inert, but rather a decisionmaker choosing what is best for its life. The assumptions are unabashedly individualistic, and the logic is strictly instrumental. Thus the goal of the model is not detailed empirical veridicality, nor the explanation of every feature of partisan socialization, but rather an accounting on the simplest possible basis for the broad outlines of what we know about parental transmission of party identification (PID) to children.

To accomplish this, first, the article sets out a stylized model of the flow of party benefits to voters in a two-party system. The expected value of these future benefits is identified with party identification. Next, a simple social structure is postulated in which parents and children have correlated but not identical positions, and these social positions influence party benefits. Thus the electoral experience of children is correlated with that of their parents, and so they can use parental experience to partially forecast their own future experience. Finally, children's optimal use of parental political experience is derived and shown to account for some central stylized facts in the partisan socialization literature.

PARTY IDENTIFICATION AND VOTING: THE ASSUMPTIONS

This section sets out a model of how voters decide. The perspective on voting is the same as that of Achen (1992). Voters choose a party or candidate when they believe that its future course of benefits exceeds that of the other alternatives. When the voters expect that a party will favor them in the future, they will be said to "identify" with that party. Of course, none of the emotional or affective content sometimes associated with identification is treated here. Those feelings are obviously quite real, but this model sets them aside in the same way and for the same reason that economics sets aside the consumer's joy at finding a perfect apple in the grocery store.

The model treats voters as Bayesian prospective decisionmakers (forward-looking and future-oriented), though with imperfect information. They use past experience to predict the future because, under the assumptions, retrospection gives the best prospective forecast (Downs, 1957). Thus at a given time, voters have a "standing decision," but they are constantly modifying it in light of events (Key, 1966).[1] The voters' use of this information is assumed to be optimal in the Bayesian sense. The model builds on closely related arguments and empirical support from Jackson (1975a, 1975b), Fiorina (1981, ch. 5), Franklin and Jackson (1983), and Franklin (1984), and follows prior Bayesian theorizing about voters by Zechman (1979) and Bartels (1983).

Within a given party alignment regime, it is assumed that a voter is offered by each party a stream of benefits (cardinal utilities) that varies randomly around a constant mean. In a two-party system, only the difference between the two offered streams matters to the vote choice.[2] Subtract the second party stream from the first, and let δ_i^* denote the mean difference for the *i*th voter, assumed constant over time (until the next realignment): Then, denoting voter *i*'s party difference in benefits at any given time *t* by u_{it}:

$$u_{it} = \delta_i^* + v_{it} \tag{1}$$

where v_{it} is normally distributed with mean zero, variance $\omega_i^2 > 0$, and is distributed independently over time. Thus benefits oscillate randomly around a fixed personal mean δ_i^* for each voter, representing a stable world between realignments in which the parties continue to stand for the same benefits over time.[3]

From this simple dynamic model of Bayesian retrospective voting across multiple elections, some of the best known findings of the voting literature were derived. For example, it follows from the assumptions that partisans will be better informed than Independents, that successful incumbents will disproportionately attract young voters to their party, and that voters with intermediate amounts of information will be the most likely to defect from their party identification (Achen, 1992).

For analytic simplicity, the model ignores another feature of the political world, Key's (1959) "secular realignment," in which realignment is continual rather than intermittent. Thus for empirical purposes, greater realism would be needed. One obvious way to model secular realignment and to meet the demands of empirical work would be to generalize the white noise assumption about v_{it}, as was proposed in Achen (1992, p. 210). Gerber and Green (1998) adopt this suggestion and take a useful step forward, replacing the white noise ARMA(0,0) disturbances with an ARMA(1,0) form in order to do statistical work.[4] Bartels (2001b) goes further, adopting a very flexible model to assess which functional forms describe voter updating. He finds that the data generally support the white noise assumption but, as he notes, this finding seems implausible.

Learning the best functional forms for statistical models of Bayesian voter updating is an important agenda item for students of public opinion. However, tracking the empirical details, consequential as they are, is not the point of this article. The goal instead is to use a stylized model to generate qualitative predictions matching the main empirical generalizations.

SOCIAL STRUCTURE AND THE PARTIES: ASSUMPTIONS

From a rational choice perspective, a question that the social-psychological literature has nearly always overlooked is the natural first one raised earlier: Why should children decide to think like their parents? The answer given here is that parent and child will often occupy similar positions in the social structure, and thus parental experience is likely to be relevant to the child's future adult life.

Proceeding in this way answers a question raised by virtually all the Bayesian theoretical models used in political science, which include "subjective prior information" as part of the voters' conceptual machinery. The voters are assumed to know something even before they acquire experience with the political system. But how? Achen (1992) proposed to interpret this prior information as parental socialization, but the rational choice foundation of that view was laid out only informally. This section makes the logic explicit and extends the findings of the earlier paper.

To formalize the intuition, a simplified social world is now constructed in which each parent and each child is characterized by a vector of social attributes, normally distributed over the population, which influence the flow of party benefits. Thus voter *i*'s position in the social structure is characterized by an *k*-dimensional (row) vector s_i, fixed over time and exogenous with respect to party benefits, whose elements may customarily be thought of as her income, social status, religious preferences, and so on.[5] This list of the voter's politically relevant personal characteristics will be called her "social position." Thus, a variable is part of the voter's social position if it affects her flow of benefits from the political parties but is causally prior to them.[6]

The party benefit differential for each voter is assumed to be a linear function of her social position. For example, a party might offer more benefits to those with higher incomes or evangelical Protestant beliefs, and the effects might be additive. Of course, party benefits are influenced by many kinds of ideological, attitudinal, and other factors not included in the voter's social position, all involved in complex causal interrelationships with party identification. Such effects are not ignored or assumed away here, but rather mathematically converted to their dependence on predetermined factors. In econometric terms, the relationship between social position and party benefits is a "reduced form," combining in the standard way the direct effects of the social variables with their indirect effects through other endogenous variables.

Recalling that δ_i^* represents voter *i*'s mean difference in party benefits, we have:

$$\delta_i^* = \alpha + s_i\beta \qquad (2)$$

where the scalar α and the vector β are parameters fixed across individuals.[7] In practice, of course, some actual voters will know, perhaps from an early age, that their party benefits will relate to their social situation or their parents' social situations in atypical ways, and they will follow a somewhat different law. But again, the goal of this article is to predict aggregate features of the socialization process using the average behavior of the population, not to model every individual. Anyone who uses ordinary regression models with their fixed coefficients has adopted the same procedure.

The intercept term α in (2) allows one party to offer greater benefits on average than the other. For example, young voters in a frozen political system might favor a party of the Left to improve their life prospects, or a party of the Right to restore peace and quiet after a tumultuous era. To simplify matters, α will be assumed known.[8] In that case, voters may as well focus on $\delta_i = \delta_i^* - \alpha$, and this paper will do so as well.

Then from Equation (2), we obviously have:

$$\delta_i = s_i\beta \tag{3}$$

For ease of exposition, δ_i will be referred to as the "party differential" in benefit streams, although that phrase more properly applies to $\delta_i + \alpha$.

Now, to determine the aggregate distribution of party benefits, assumptions are made about the distribution of the social position vector s_i across the population of voters. It is taken to be multivariate normally distributed in each generation. Thus if parental social position is denoted by s_p, that of their children by s_c, and if the stacked vector of parents and child is $s = (s_p\ s_c)$, then it is assumed that:

$$s \sim N(\mu, \Sigma) \tag{4}$$

where $\mu = (\mu_1\ \mu_2)$ is the $2k$-dimensional mean vector and Σ is the joint variance–covariance matrix, assumed positive definite. Thus in particular, the upper-left $k \times k$ submatrix Σ_1 is the variance–covariance matrix for the parents, and the analogous lower-right submatrix Σ_2 that of the children. The positive–definiteness assumption means that no dimension of the child's social position can be perfectly forecast knowing only the parents' social position, and vice versa. Since the scales are arbitrary, one may set $\mu_1 = \mu_2 = 0$ and $\Sigma_1 = \Sigma_2 = \Sigma_s$.

The square off diagonal submatrices $\Sigma_{12} = \Sigma_{21}'$ give the covariance matrix between s_p and s_c. They will be assumed positive–definite; this is the multivariate version of the assumption that parents and children have positively correlated but not identical social positions.[9]

These assumptions about s_i imply from Equation (3) that δ_i has a normal distribution across the population. Moreover, since cardinal utilities are unique only up to a linear transformation, one may scale δ_i to have unit variance, so that $\beta'\Sigma_s\beta = 1$. Thus δ_i is standard normal.

Now the multivariate normality assumption of Equation (4) implies that parents transmit their social position, s_p, to their children, s_c, according to a linear law. That is, each element of the child's social position is a linear function of all the elements of its parents' social position:

$$s_c = s_p R + w \tag{5}$$

where Equation (5) is a “seemingly unrelated regressions” formulation with identical independent variables in each regression, the matrix of regression coefficients is $R = \Sigma_s^{-1}\Sigma_{12}$ by the usual least squares formulas, and w is a multivariate normally distributed disturbance with mean zero and variance matrix $\Sigma_s - R'\Sigma_s R$.[10] The main diagonal elements of R will be assumed strictly positive. That is, parents' and children's corresponding elements in their social positions are positively correlated.

One may get a little intuition about this relationship by noting that $R'R$ is just the squared correlation matrix between elements of s_p and elements of s_c. Thus for example, in the one-dimensional case, s_p and s_c might represent just social class. Then R in Equation (5) would be a single regression coefficient, and it would equal the correlation coefficient between the social class of parents and their children.[11]

PARENTS, CHILDREN, AND THE PARTIES

To see how social structure and party benefits are transmitted from parent to child, begin with Equation (3). There, the voter's party benefit differential depends on her social position. Applying this to the case of the child gives:

$$\delta_c = s_c \beta \tag{6}$$

Next, substituting for s_c from the social transmission Equation (5) gives the relation between the child's party benefits and parental social position:

$$\delta_c = (s_p R + w)\beta \tag{7}$$

Now, voters are assumed to be generally unaware of all this social structure. They are ignorant of the social transmission process, their own and their parents' social position vectors, and the true values of their own and their parents' expected benefit streams. Thus the child cannot use Equation (7) directly to forecast future benefits.[12] However, it is possible for the child to learn how parental experience with the parties, δ_p, correlates with the child's future benefits, δ_c. Since the parents' benefits depend on their social position according to $\delta_p = s_p\beta$, the covariance between δ_p and δ_c is, using Equations (5) and (7):

$$E(\delta_p'\delta_c) = E(\beta' s_p' s_p R\beta) + E(\beta' s_p' w\beta) \tag{8}$$

But β is fixed and s_p and w are uncorrelated from Equation (5), so that the last term is zero.

Now one may solve for the correlation between mean parental party benefits δ_p and their children's future benefits δ_c. Call this correlation ρ.[13] Then, using the previous equation, along with the fact that the variance–covariance matrix of s_p is Σ_s:

$$\begin{aligned} cov(\delta_p, \delta_c) &\equiv \rho \\ &= \beta' \Sigma_s R \beta \end{aligned} \tag{9}$$

It is easily shown that under the assumptions of this article, $\rho > 0$, that is, children tend to have experiences with the political parties similar to those of their parents. However, the correlation is imperfect and children cannot simply copy their parents: $\rho < 1$.[14]

Now we may spell out the predictive relationship between the child's experience with the parties and that of its parents. Using Equation (9), since all variables are normally distributed, we have:

$$\delta_c = \rho\delta_p + w\beta \tag{10}$$

This is a decomposition of the voter's expected benefit stream from the parties, δ_c. The first part is her expected benefit stream due to her partial inheritance of the social position of her parents and thus her correlation with the party benefits δ_p they experience during their lifetime. The second part is that due to her unanticipated difference in social position from that of her parents.

It may be useful here to summarize all the basic assumptions that led to Equation (10). First, parent and child positions in the social structure are normally distributed and hence linearly related with an error term. Second, for both the child and the parents, mean experience with the parties is linearly related to their social position. Third, for both children and parents, each period's experience with the parties equals their mean experience plus a normally distributed error term. Hence by standard Bayesian logic, if children know their parents' PID, they can forecast their own by adopting their parents' PID but then regressing it somewhat toward 0, that is, toward lesser partisanship.[15] That is the content of Equation (10).

As already mentioned, the voters are assumed to know very little. However, they do know something. Concretely, they know only that if their parents disliked Republicans, then it is likely, but not certain, that their experience in adulthood with the GOP will be negative, too. This substantive assumption is embodied in the model's assumption that the voters know Equation (10), and they know ρ. These two bits of statistical knowledge make it possible to use parental experience as an guide on entering political adulthood. Of course, the voters need not literally know the model and the value of its parameter for them to behave in the qualitative manner it describes, anymore than tennis players returning serve need to know the differential equations describing the flight of tennis balls. To validate the assumption, voters need only treat parental PID as a helpful but not infallible guide for themselves.

As it stands, however, Equation (10) is useless to the child. It relates the true long-term PIDs of parents and children to each other, neither of which will ever be known. Parents do not live forever, and even if they did, party alignments do not persist indefinitely. Hence children must make do with what their parents have learned thus far.

The child copes with this limitation in the following way. At time 0, the voter is a young adult just entering the political system and considering her future within it. For many, this step occurs when they first become eligible to vote, though of course it may occur somewhat earlier or later as well.[16] She has no experience with the party system herself and hence no direct evidence of its mean benefit difference for her, δ_c. However, she knows Equation (10) plus her parents' mean flow of benefits since the last realignment, namely $\bar{u}_{p0}$.The parents' current estimate of their mean benefits (their current PID) may be written as:

$$\bar{u}_{p0} = \delta_p + \varepsilon_{p0} \tag{11}$$

where $\varepsilon_{p0} \sim N(0,\sigma^2_{p0})$.[17] Note that the normality of ε_{p0} follows from that of u_{it} in Equation (1).

With this structure, a new voter (the "child" in the socialization literature, but actually a teenager) is ready to make a rational choice of her party identification. To make the logic explicit, note that since $\bar{u}_{p0}$ is an estimate of δ_p, the new voter may substitute it for δ_p on the right-hand side of equation (9) to produce a forecast for her own δ_c. This is a bivariate regression setup with errors in the independent variable. Since all variables are normally distributed, the optimal forecast is given by ordinary least squares formulas. The initial party identification is just the forecast of δ_c at time 0. Denoting it by $\hat{\delta}_{c0}$, the new voter is best off selecting the following value for her initial party identification:

$$\hat{\delta}_{c0} = \rho\lambda\bar{u}_{p0} \tag{12}$$

where by the usual formula for attenuation of the slope, $\lambda = 1/(1 + \sigma^2_{p0})$, which is (true variance of δ_p)/(true variance + error variance). The parameter λ measures the reliability of party benefits: When the parties dependably deliver approximately the same benefit level each period, λ nears 1.0. When benefits are highly erratic and undependable, λ approaches 0. In the latter case, parents will be less sure about their PID, and their children should rely less on their parents, as Equation (12) implies.

The new voter's forecast variance for her party benefits is easily seen to be $1 - \rho^2\lambda^2(1 + \sigma^2_{p0}) = 1 - \rho^2\lambda$. Put another way, at time 0 the child is not entirely certain about her future experience with the parties, but she has expectations that allow for some uncertainty, expressed as a statistical distribution. The posterior distribution for her mean future benefit stream is distributed as $N(\hat{\delta}_{c0}, 1 - \rho^2\lambda)$.

With this structure, several different concepts of party identification can be given explicit definition. Kramer (1987) argued that too many important political science concepts have only an intuitive meaning; "party identification" is an outstanding example. Traditionally the concept is defined only informally—as an "attitude." In practice, it is most often identified with the answer to an opinion survey question, and the inevitable debates about wording and cross–cultural equivalence have ensued.

In this article, the term "current party identification" of voter i will refer precisely to $\hat{\delta}_{i0}$ (and its updated estimates in subsequent periods, $\hat{\delta}_{in}$). The constant δ_i, to which $\hat{\delta}_{in}$ tends, will be called the "long–term party identification." This distinction makes it clear how a voter can have simultaneously a stable PID to which she tends to return and also another PID that she is constantly revising and updating in response to contemporary events. It also suggests that in two–party systems, PID can be thought of as a three-dimensional quantity—estimated mean benefit level from Democrats, estimated mean benefit level from Republicans, and the variance or uncertainty in these estimates (compare Weisberg, 1980).

The Michigan survey response is yet another quantity, reflecting current PID plus the usual survey noise. (Dramatic examples of how differently people interpret the usual questionnaire items about partisanship appear in Flanigan, Rahn, and Zingale, 1989.) For purposes of verification, students of voting have little alternative to using the survey responses about partisanship as proxies for the theoretical quantities, and this article will follow in that tradition. Happily, PID is among our best measured and most reliable concepts, and so the problem is minimized here. But it bears emphasizing that the formal propositions of this article refer to the variables as defined by the theory, not to less reliable survey research operationalizations of precise concepts such as "prospective evaluations" or "perceived party benefits." For understandable reasons, surveys rarely tie their questions to the strict meaning theory gives them, and respondents often respond with something else entirely. For that reason, when such variables seem to contradict the implications of a formal theory but accuracy of measurement is not assessed, survey noise is always a very credible alternative explanation (compare Gerber and Green, 1998, pp. 798–801).

THE MAIN EMPIRICAL IMPLICATIONS

With the model of the previous section in hand, several propositions familiar from the political socialization literature emerge immediately, as do several other empirical implications.[18] Before turning to them, however, a preliminary result is needed.

The first proposition compares the initial PID of new voters to that of their parents. To make that comparison, the parents' current PID is needed. The same calculations used for the new voter may be applied to the parents to give the parents' current best estimate:

$$\hat{\delta}_{p0} = \lambda \bar{u}_{p0} \tag{13}$$

with posterior variance $1 - \lambda$.[19]

With this result, the first proposition can be stated:

Proposition 1. New voters' current PIDs are initially positively correlated with that of their parents but more centrist: $cov(\hat{\delta}_p, \hat{\delta}_{c0}) > 0$ and $|\hat{\delta}_p| > |\hat{\delta}_{c0}|$.
Proof. Using the above Equations (12) and (13), and $0 < \rho < 1$: $|\hat{\delta}_{p0}| = |\lambda \bar{u}_{p0}| > \rho\,|\lambda \bar{u}_{p0}| = |\rho\lambda \bar{u}_{p0}| = \hat{\delta}_{c0}|$.

Thus young voters tend to have their parents' party identifications, but as a group, they are Independents more often than their elders. Evidence for this proposition in the United States appears in Jennings and Niemi (1981, pp. 90–91, 153). Note again that "centrist" here means "toward the middle of the distribution": new voters regress toward the population average of the party differentials, which need not be 0. Thus, in an era of Democratic PID advantage, initial PIDs will be a compromise between parental PID and a mild Democratic tendency, as the Jennings-Niemi data show.

Notice also that Proposition 1 refers to a rather abstract class of people—young voters who know absolutely nothing of the political world except their parents' PID. The Proposition implies that among these idealized young voters just at the exact moment before they begin to learn on their own, some might be Independents, but absolutely none of them will have the opposite partisanship from their parents. Testing such a claim with survey data is obviously impossible, and not of much importance in any case.

In what sense, then, might the Proposition be validated? It is important to remember that surveys of young voters do not reach them at the moment they enter the political system in a state of complete innocence apart from parental preferences, if indeed such a state exists. Instead, actual young voters have often moved beyond the Proposition's semifictional time 0, meaning that they possess varying amounts of experience and knowledge of the political system. In that case, if Proposition 1 is correct, then a corollary can be proved (though at greater length and with less transparency than the Proposition). On average, new voters will have begun to move toward their mature PID, but from a starting point of Independence or agreement with parental PID. Thus, most young voters will have their parents' PID or none at all, either because they have not yet given much thought to politics or because they are finding that their adult political preferences match that of their parents. However, a minority of new voters will have acquired enough initial experience to deviate from their parents and adopt the opposing partisanship. The corollary predicts some strict opposition to parental PID, but not much. That is, of course, just what the data show.

Thus the Proposition's qualitative forecasts—strong agreement between parent and child, but more Independents among children—are matched by the data, even though its precise predictions are untestable. A corollary to the Proposition also makes the same qualitative predictions, is directly testable, and is confirmed. It is in these senses that the Proposition is confirmed. Similar remarks refer to all the succeeding Propositions in which new voters play a role.

Proposition 2. New voters' current PIDs are initially more labile (higher subjective variance of estimate) than their parents.
Proof. The new voters' variance is $1 - \rho^2\lambda$; that of parents is $1 - \lambda$; $\lambda > 0$; the conclusion follows.

For suggestive evidence in the American context, see Jennings and Niemi (1981: 50–51), where strength of PID is shown to be lower for young adults than for their parents.

Proposition 3. New voters with more labile parents have more labile PIDs; they will also have more centrist PIDs than other new voters whose parents have the same mean experience with the parties: if p and c denote one parent–child combination and p' and c' another, then $\bar{u}_{p0} = \bar{u}_{p'0}$ and $\sigma_{p0} > \sigma_{p'0}$ implies both $1 - \rho^2\lambda_p > 1 - \rho^2\lambda_{p'}$ and $|\hat{\delta}_{c0}| < \hat{\delta}_{c'0}|$.
Proof. By definition, $\lambda_p = 1/(1 + \sigma_{p0})$, and similarly for $\lambda_{p'}$. Hence $\sigma_{p0} > \sigma_{p'0}$ implies $\lambda_p < \lambda_{p'}$. In combination with $\rho > 0$, this clearly implies the inequality on the posterior variances. Finally, note that since $|\hat{\delta}_{c0}| = \rho\lambda_p\bar{u}_{p0}$ and similarly for $|\hat{\delta}_{c'0}|$, the inequality on the λ's immediately implies, in combination with $\bar{u}_{p0} = \bar{u}_{p'0}$ and ρ, λ_p, $\lambda_{p'} > 0$, that $|\hat{\delta}_{c0}| < |\hat{\delta}_{c'0}|$.

There does not seem to be a direct assessment of Proposition 3 in the literature. Rough tests appear in Campbell et al. (1960/1980, p. 147), who use respondents' reports of their parents to show that, controlled for parental PID, those raised in less "politically active" homes have more centrist PIDs. Similarly, Jennings and Niemi (1981, pp. 86–87) use parents' and children's own reports to demonstrate that the children of politically poorly informed parents tend to be poorly informed.

Proposition 4. All else equal, the more limited or more erratic the political experience of the parental generation (smaller λ), the more centrist (Independent) the initial PIDs of young voters will be.
Proof. Obvious from Equation (12). See the discussion there for additional remarks and interpretation.

This is the standard explanation for the low levels of partisanship found in Germany, France, and Italy in the immediate decades after World War II (Sears, 1975, pp. 120–121). Fascist interludes and regime change make parental political experience thin, less reliable, and thus less relevant to new generations. Hence the young tend to disregard their parents.[20]

Proposition 5. All else equal, the greater the changes in party policy across generations (smaller ρ, meaning that similar people in different generations receive dissimilar benefits), the more centrist (Independent) the initial PIDs of young voters will be.
Proof. Again obvious from Equation (12).

Any factor that reduces the correlation between party benefits to similar people in different generations, such as social mobility, will have the same effect:

Corollary. All else equal, greater social mobility (an attentuated main diagonal element in R) induces more centrist (Independent) PIDs among young voters.
Proof. The relationship between the social position of parents and that of children is given by the matrix R in Equation (5). When it is a scalar, Equation (9) obviously implies that a smaller R produces a smaller ρ, so that by Proposition 5, the result is proved. With a bit more work using partitioned matrices, it may be shown that the same is true if one main diagonal element of R is reduced in magnitude.

To my knowledge, the hypotheses just stated in the Proposition and its Corollary have never been investigated.

In a different interpretation, Proposition 5 also explains why teenagers ignore parental advice about haircuts and popular music. On topics like these, parental experience is a very poor predictor of what the next generation will find attractive. Those aspects of social position that predict such preferences are poorly correlated over generations, inducing a small ρ. Hence teenagers sensibly pay their parents no mind.

Proposition 6. Large party benefits at a given time period have effects that are transmitted across generations but die out over time.

Proof. A one-time shock g in the parental generation raises their experience of party benefits when their children come of age, $\bar{u}_{p0}$, by the amount g/n, where n is the number of elections experienced by the parents.[21] By Equation (12), the extra quantity $\rho\lambda g/n$ is then transmitted to their children. The succeeding generation will then be transmitted the extra amount $\rho^2\lambda^2 g/n^2$, and so forth. Since both ρ and λ are strictly bounded between zero and one, the Proposition follows.

The literature on realignments and their subsequent decay provides broad support for this proposition. Beck (1979) in particular discusses the cyclical pattern of American electoral politics, in which the initial shock of a realignment gradually wears off and new alignments become possible. As he notes, a simple possible explanation for the cycle is that the generations pass and children forget their grandparents' lives, making subsequent generations available for recruitment to new party structures. Unfortunately, careful empirical testing of this proposition appears nearly impossible.

For the final result, an additional variable is needed. This variable is "affective warmth" or "closeness" between parent and child. Denote it by a_{pc}. Experience suggests that this variable is not much related to demographic similarities between parent and child such as gender, occupation, or religious beliefs. Some parent–child pairs love each other deeply, most care a great deal about each other, and a few are oil and water, but in all these cases it is rare for there to be much impact on the child's choice of occupation or lifestyle.

For the purposes of this article, no soundings of these depths is required. We need only assume that once we have predicted the child's social position s_c from the parental social position s_p, their closeness a_{pc} adds no explanatory power:

$$E(s_c \mid s_p, a_{pc}) = E(s_c \mid s_p) \tag{14}$$

This is obviously only an approximation, but probably a good one. For example, children with lawyer parents are more likely to wind up as attorneys than other children: That correlation is embedded in the equation. What the equation adds is an assumption that the correlation does not increase when the children love their attorney parents deeply. Indeed, most well-loved children of attorneys are not attorneys.

The assumption in Equation (14) has a strong consequence very much in contrast to the traditional social-psychological theories of socialization. Because Equation (14) is speculative, this implication will be labeled a conjecture:

Conjecture. Affective closeness between parents and children has no effects on the socialization of partisanship. In particular, the children of close relationships are no more likely to share their parents' PID than are the children of less close relationships who are similarly placed in the social structure.

Proof. By Equation (3), party benefits depend only on social position, so that parents and children have correlated party benefits only because their social positions are correlated. By the assumption in Equation (14), affective closeness adds nothing additional to the prediction of the child's social position or future party benefits. Thus in Equations (10) and (12), this new variable will be ignored. Hence the previous logic and results hold in all detail whether or not the child is close to the parents, as the Conjecture states.

Closeness to parents as a cause of political socialization has been studied from time to time. Its effects have usually turned out to be small or erratic (see the review in Sears, 1975, p. 126), much to the researchers' surprise. If the assumption in Equation (14) holds, however, the weak effects are natural. The warmth of family relationships does not keep family members from ignoring useless advice.[22]

This result, so surprising from the social-psychological perspective, is one more reason why the rational choice viewpoint is helpful. Treating the teenage child as a thinking being and asking *why* the child should adopt the parents' views puts the emphasis on those parental preferences and relationships that are likely to be helpful to the child in future. In turn, that sorts out the factors likely to be influential in socialization from those that are not. The older literature tended to treat the child as a much-loved vessel into which parents poured their values. In consequence, it relied on a different set of explantory variables, more affective and less instrumental. Thus a careful empirical test of the importance of parent–child closeness in partisan socialization would help sort out the relative contributions of rational choice and social-psychological perspectives to the understanding of political socialization.

CONCLUSION AND DISCUSSION

This article has set out a very simple rational choice model for parental socialization of children's party identification. The model is closely integrated with previous Bayesian models of party identification, the effects of political campaigns, and vote choice. The various Bayesian models imply the key qualitative findings of the social-psychological school.

With respect to partisan socialization in particular, this article assumed that children will have adult social positions correlated with that of their parents and that social positions are related to benefits received from political parties (party identification). Hence new voters behaving according to standard Bayesian logic can learn something of their own future benefits by knowing the experience of their parents and then as they get more experience, they can update their PID. Several standard findings of the behavioral literature were derived from this framework, including the tendency of young voters to have the same identification as their parents, the greater proportion of Independents among young voters, the low levels of partisanship found in political systems after dictatorial interludes, and the tendency of party alignments to decay. Thus by deriving the central empirical generalizations in the literature, Bayesian voter theory both explains why they typically hold and suggests conditions under which the axioms might fail and the generalizations not apply.

This theoretical success raises a puzzle. For 50 years, a decent respect for social science evidence (or just walking a precinct) has compelled social scientists to the view that Bayesian updating is not a good empirical description of voter decisionmaking. Yet without any need for ad hockeries or "adjustment" of assumptions, the simplest possible Bayesian models generate strong, nonobvious, but well-verified logical implications across the entire field of voting behavior, including the study of socialization. How can this be so?

Three forces seem to be at work. First, the great majority of the evidence cited in this article refers to studies of population or subpopulation averages. Often these averages are the only socially relevant outcomes. But then, as long as the average voter responds correctly, the data will look as though everyone did. The fact that the model is rather poor at the individual level may make almost no difference for purposes of social explanation. A large group covers a multitude of inferential sins.

Second, the voter is rarely alone when making decisions. Interest groups, politically aware family members and friends, and the media all spend time telling her what she should be thinking, and many of them have the resources to update their estimates with great care. If the voter knows only enough to listen to people whose basic political interests resemble her own, she may once again behave as if she were a Bayesian updater even though in fact she knows none of the relevant information. Thus individualist logic often nicely mimics more complex social processes while bypassing the formidable demands of modeling them accurately.

Third and finally, the fundamental Bayesian logic is qualitatively robust against certain kinds of errors. Consider the result sometimes found in decisionmaking studies, in which the voter is a "conservative Bayesian": for example, she might make use of her parental PID, but not as strongly as the social structure demands. Thus in Equation (12), she may use "too small" a weighting factor ρ^*, say, where $\rho^* < \rho$. Modifying Equation (12) accordingly then gives:

$$\hat{\delta}_{c0} = \rho^* \lambda \bar{u}_{p0} \tag{15}$$

The above equation describes an inept Bayesian. However, in a slightly different world in which party benefits were less correlated across generations (that is, where ρ^* were the true value rather than a mistake), the equation would describe a flawless Bayesian. All the same qualitative findings given in the Propositions would hold either way. Since the underlying parameters are unknown and perhaps unknowable, the distinction is irrelevant. There is no way to tell which world we are in, and it makes no qualitative difference. Since qualitative predictions and not quantitative data fitting is the point of the model, this particular failure of rationality, which looms so large in the social-psychological literature, has essentially no empirical consequences.[23] In short, there is reason to expect that, given appropriate auxiliary assumptions, a Bayesian approach to formal modeling of voter choice will often work very well, especially for aggregate predictions, in spite of somewhat inaccurate assumptions.

Acknowledgments. Prepared for presentation at the Parties and Partisanship Conference, Vanderbilt University, October 25–27, 2001. I am indebted to participants in that conference for helpful criticism and suggestions, as well as to the Center for the Study of Democratic Politics at Princeton University and the Department of Political Science at the University of Michigan, for financial support. I also thank Larry Bartels, Stanley Feldman, William Flanigan, John Freeman, Fred Greenstein, John E. Jackson, Robert Luskin, Michael MacKuen, Douglas Rivers, J. Merrill Shanks, James Stimson, Margaret Trevor, and John Zaller for much useful assistance. My gratitude as well to the unusually helpful editors and to two anonymous reviewers, in particular Reviewer B, whose painstaking reading and trenchant critical intelligence substantially improved the final version. Not having taken all the advice I received, I have the responsibility for remaining errors.

NOTES

1. Bayesian models often allow the voter to incorporate current campaign information into PID as well. Voters clearly do so, an effect visible in survey responses. This short-term volatility is irrelevant to the study of socialization, which focuses on long-term effects. However, the distinction is worth remembering when assessing socialization models using a survey like the American National Election Study (NES), which is carried out during election campaigns.
2. The extension to the more usual case of multiple parties is essentially straightforward, although the arithmetic becomes heavier, since multidimensional measures of PID are required.
3. Throughout this article, the model will be interpreted as applying to national politics, where nearly all partisan socialization research has been focused. However, a voter might have different party benefit streams from national, state, and local politics, and thus have party identifications that vary across levels. Canada is the outstanding example.
4. Oddly, Gerber and Green (1998) treat their statistically oriented model as competing with the stylized setup in Achen (1992), where the white noise model was explicitly set out as a "special case" and an example of "crude models" (pp. 202, 209) built for theoretical clout. No position was taken there about the correct empirical ARMA specification, and statistical exploration like Gerber and Green's was encouraged.

5. Since women are the majority, I will refer to voters most often as "she."
6. Naturally, "social position" is defined relative to a time, place, and sociopolitical system. It is easy to think of situations in which the demographic and occupational variables routinely treated as exogenous in our regression equations would instead be driven by politics and would therefore be endogenous. Thus different political systems will have different variables included in voter social positions.
7. Readers who would prefer a disturbance term in this equation may simply interpret the last element of s_i as the disturbance and set the last element of β to one. Under the usual independence conditions on disturbances, this stochastic formulation is actually less general than the seemingly more restrictive deterministic setup in Equation (2).
8. This assumption would require American voters to know how often the Democrats or the Republicans win (since electoral outcomes are determined by the balance of party benefits α). For greater realism, voter uncertainty about this parameter could be incorporated into the model in the standard Bayesian manner, although at the cost of considerable clutter and no important changes in the substantive findings.
9. Note that there is exactly one parent per child; "parents" are treated here as if they were a single person or an average. This is an obvious oversimplification, but it seems hopeless to model in full detail the tag-team socialization of contemporary American children.
10. The inverse Σ_s^{-1} required to define R exists because Σ is positive definite and hence of full rank $2k$. Thus its principal diagonal square submatrices such as Σ_s are also of full rank and have an inverse.
11. Notice that discrete variables like gender are not normally distributed, and minor modifications to the model would be required to incorporate their political effects.
12. The argument of this article would be little changed if children were aware of their parents' (multidimensional) social position and knew how all its components combined to generate party identification. Essentially, children would then have several noisy predictors of their own partisanship rather than just parental PID, but the fundamental Bayesian logic would be unaltered, and the qualitative results of this paper would remain unaltered.
13. Note that since δ_p and δ_c are both standard normal, their covariance ρ is also their correlation coefficient.
14. The strict non-negativity of ρ holds because $\rho = \beta'\Sigma_s R\beta = \beta'\Sigma_{12}\beta$, and Σ_{12} is positive–definite by assumption, so that $\rho > 0$. On the other hand, $\rho < 1$. To see that, let the stacked $2k$-dimensional vector $\beta^{*\prime} = (\beta' - \beta')$ and consider the quadratic form $\beta^{*\prime}\Sigma\beta^*$. Since Σ is positive–definite, the latter term is strictly positive. Simple partitioned matrix multiplication gives its value as $2\beta'\Sigma_s\beta - 2\beta'\Sigma_{12}\beta = 2 - 2\rho$ by the discussion below Equation (3) and by the previous paragraph. But since $2 - 2\rho$ must be strictly positive, then $\rho < 1$.
15. More precisely, in forming their initial PID, children should regress their parents' partisanship toward the average partisanship α of the population. This may not be near the 0 (neutral) point between the parties in some time periods. See the discussion earlier of Equations (2) and (3).
16. As one reviewer correctly pointed out, the 10-year old quoted at the beginning of the article is surely outside the model, with her short-term fate tied more closely to her parents than is the case for the new voters to which the model is directed. However, the logic of the article would need only a slight extension to cover that case. It would imply greater parent–child agreement among 10-year olds than among 18-year olds. The literature cited at the beginning of this article will confirm that prediction, as will anyone who has lived with children of both ages.
17. The parents' current estimate of party benefits actually includes some prior information based on the experience of *their* parents (the child's grandparents), which complicates the logic slightly without affecting the conclusions. For simplicity, that feature of parental PID is ignored here. The model's full logic is given in Achen (1992).

18. The first three were set out without proof in Achen (1992, p. 205).
19. Note that parental PIDs are not just the raw average of their experiences, $\bar{u}_{p0}$. Instead, their PIDs are regressed to the population mean via the parameter λ, due to the existence of prior information about the distribution of δ_i. The same logic explains the appearance of λ in the equation for the child's initial PID, Equation (12).
20. Strictly speaking, the proposition relates only to changes in λ, which is driven by the variability in the voter's party benefits in Equation (1). Regime changes and undemocratic intervals would modify other aspects of the model as well, but the argument in these more dramatic cases would be similar.
21. Actually the divisor is slightly larger, due to the prior transmitted by the grandparents' generation (Achen, 1992), but that makes no difference here.
22. Of course, it is not hard to think of situations in which the effects would be stronger. The point is that a variable should influence the new voter's partisanship only when it predicts her future relationship to the political system.
23. Of course, precisely the same argument applies to a "liberal Bayesian" with $\rho^* > \rho$.

REFERENCES

Achen, Christopher H. (1992). Social psychology, demographic variables, and linear regression: breaking the iron triangle in voting research. *Political Behavior* 14, 3 (September): 195–211.

Bartels, Larry M. (1993). Messages received. *American Political Science Review* 41, 2 (April): 641–674.

Bartels, Larry M. (2001a). A generational model of political learning. Presented at the annual meeting of the American Political Science Association, San Francisco, September.

Bartels, Larry M. (2001b). Beyond the running tally. Presented at the Parties and Partisanship Conference, Vanderbilt University, October 25–27.

Campbell, Angus, Converse, Philip E., Miller, Warren E., and Stokes, Donald E. (1980). *The American Voter* (Midway Reprint). Chicago: University of Chicago Press. (Original work published 1960)

Converse, Philip E. (1964). The nature of belief systems in mass publics. In David E. Apter (ed.), *Ideology and Discontent*. Glencoe, IL: Free Press.

Converse, Philip E. (1969). Of time and partisan stability. *Comparative Political Studies* 2: 139 71.

Converse, Philip E., and Pierce, Roy (1985). Measuring partisanship. *Political Methodology* 11, 3–4: 143–166.

Converse, Philip E., and Pierce, Roy (1986). *Political Representation in France.* Cambridge, MA: Harvard University Press.

Downs, Anthony (1957). *An Economic Theory of Democracy.* New York: Harper and Row.

Easton, David, and Dennis, Jack (1969). *Children in the Political System.* New York: McGraw-Hill.

Fiorina, Morris (1981). *Retrospective Voting in American Elections.* New Haven: Yale University Press.

Flanigan, William H., Rahn, Wendy M., and Zingale, Nancy (1989). Political parties as objects of identification and orientation. Presented at the Western Political Science annual meeting, Salt Lake City, March.

Franklin, Charles H. (1984). Issue preferences, socialization, and the evolution of party identification." *American Journal of Political Science* 28, 3 (August): 459–478.
Franklin, Charles H., and Jackson, John E. (1983). The dynamics of party identification. *American Political Science Review* 77: 957–973.
Gerber, Alan, and Green, Donald P. (1998). Rational learning and partisan attitudes. *American Journal of Political Science* 42, 3 (July): 794–818.
Green, Donald P., Palmquist, Bradley, and Schickler, Eric (2002). *Partisan Hearts and Minds.* New Haven: Yale University Press.
Greenstein, Fred I. (1969). *Children and Politics* (revised edition). New Haven: Yale University Press.
Hess, Robert D., and Torney, Judith V. (1968). *The Development of Political Attitudes in Children.* Garden City, NY: Anchor.
Jackson, John E. (1975a). Issues, party choices and presidential votes. *American Journal of Political Science* 19: 161–185.
Jackson, John E. (1975b). Issues and party alignment. In Louis Maisel and Paul M. Sacks (eds.), *The Future of Political Parties*, pp. 101–123. Beverly Hills: Sage.
Jennings, M. Kent, and Niemi, Richard G. (1981). *Generations and Politics.* Princeton: Princeton University Press.
Johnston, Richard (1988). Paper presented at the 1988 meeting of the American Political Science Association.
Kershaw, Ian (1983). *Popular Opinion and Political Dissent in the Third Reich.* Oxford: Oxford University Press.
Key, V. O. (1959). Secular realignment and the party system. *Journal of Politics* 21: 198–210.
Key, V. O. (1966). *The Res ponsible Electorate.* New York: Vintage.
Kramer, Gerald (1988). Political science as science." In Herb Weisberg, ed. *Political Science: The Science of Politics*, pp. 11–23. New York: Agathon.
Miller, Warren E., and Shanks, J. Merrill (1996). *The New American Voter.* Cambridge, MA: Harvard University Press.
Sears, David O. (1975). Political Socialization. In Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science* (vol. 2), pp. 93–136. Reading, MA: Addison-Wesley.
Shively, W. Phillips (1980). The nature of party identification: a review of recent developments. In John C. Pierce and John L. Sullivan (eds.), *The Electorate Reconsidered*, pp. 219–236. Beverly Hills: Sage.
Trevor, Margaret C. (1999). Political socialization, party identification, and the gender gap. *Public Opinion Quarterly* 63, 1 (Spring): 62–89.
Weisberg, Herbert F. (1980). A multidimensional conceptualization of party identification. *Political Behavior* 2: 33–60.
Zechman, Martin J. (1979). Dynamic models of the voter's decision calculus. *Public Choice* 34, 3–4: 297–315.