Theor Chem Acc (2006) 115: 370–378
DOI 10.1007/s00214-006-0121-5

REGULAR ARTICLE

Paul W. Ayers

Information theory, the shape function, and the Hirshfeld atom

Received: 1 November 2003 / Accepted: 7 November 2005 / Published online: 17 February 2006

Abstract Following the work of Nalewajski and Parr, there has been a surge of interest in the use of information theory to describe chemical bonding and chemical reactions. However, the measure of "information" used by Nalewajski and Parr is not any of the usual conventional entropies, chiefly because the electron density is not normalized to one. The consequences of this are discussed, and a solution is constructed using the shape function and an "entropy of mixing" term. The same revision, however, cannot be made when if the Tsallis entropy, instead of the Shannon form, is used. This serves to emphasize that the Hirshfeld atom is a very specific result, associated only with logarithmic measures of information. A less specific derivation due to Nalewajski provides one resolution to this quandary; this derivation is analyzed in detail.

1 Introduction

Recently, there has been a surge of interest in the Hirshfeld or stockholder partitioning of atoms into their composing fragments [1]. In particular, the Hirshfeld partitioning has been found to be particularly effective in studying the chemical reactivity of molecules [2–13] and has been used to gain insights, at both a fundamental and a practical level, into chemical bonding and molecular structure [14–21]. Much of this renewed interest may be traced to the work of Nalewajski and Parr [17,22] which asserts that the electron density of the Hirshfeld atoms composing an M-atom molecule can be derived by minimizing the loss of information in molecule formation, namely,

$$I\left[\left\{\rho_A(\boldsymbol{r})\right\}_{A=1}^{M}\middle|\left\{\rho_A^0(\boldsymbol{r})\right\}_{A=1}^{M}\right]=\sum_{A=1}^{M}\int\rho_A(\boldsymbol{r})\ln\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\mathrm{d}\boldsymbol{r} \quad (1)$$

subject to the constraint that the sum of the atomic densities equals the density of the entire molecule,

$$\rho_{\mathrm{mol}}(\boldsymbol{r})\equiv\sum_{A=1}^{M}\rho_A(\boldsymbol{r}). \quad (2)$$

In Eq. (1), $\left\{\rho_A^0(\boldsymbol{r})\right\}_{A=1}^{M}$ are understood to be "reference" densities, to which the atoms in the molecule are compared. In the original version of the theory, the reference densities are taken to be the densities of the isolated atoms, and so the density of the Hirshfeld atom measures the change in an atom's electron density associated with molecule formation, measured relative to the separated atom limit [22]. It is not surprising, then, that one can use the Hirshfeld atom to measure both the proclivity of a site toward chemical reactions [2–12] and the way electrons "flow" throughout a molecule [2,17].

The "information loss" criterion is not the usual language used by chemists when defining atoms in molecules. The chemists' conception of atoms in molecules is largely based on transferability. Through experimental or theoretical studies, we qualify the properties of atoms and functional groups (e.g., a nitro substituent reduces the reactivity of benzene) and use this knowledge to infer the chemical behavior of other similar systems (a nitro substituent will also reduce the reactivity of naphthalene). Similarly, using our knowledge of where peaks associated with aliphatic, acylic, and aromatic protons are found in an NMR spectrum, we can predict the NMR spectrum of a newly synthesized molecule.

The essential goal of all atom-in-molecule theories, then, is to allow us to leverage what is well-known about atoms and functional groups in one setting into a deeper understanding of less familiar molecules and molecular phenomena. More precisely, we want to maximize the transferability of the atoms between the "reference" molecule and the "target" molecule. [10] In an ideal case, the reference system is fully quantified and we wish to ensure that we can apply most – ideally all – of the information we know about the reference system to the unknown molecule. That is, we want to minimize the loss of information: we want to ensure that—to the greatest extent possible—the "information" we

---

P. W. Ayers
Department of Chemistry, McMaster University,
Hamilton, Ontario, L8S 4M1 Canada
E-mail: ayers@mcmaster.ca

know about the "reference" molecule will also be applicable in the new system we are yet to quantify [22]. Equations (1) and (2) are the mathematical statement of this idea: given that atoms computed from well-understood reference systems, minimize the loss of information subject to whatever constraints [in this simple case, only Eq. (2)], are implied by what we already know about the molecule that our studies are targeting.

2 Information theory

2.1 Properties of information

A subtle point – one these authors did not recognize in their earlier work – arises with regard to Eq. (1). In order to qualify as a measure of the "information change" relative to a reference, the form in Eq. (1) must satisfy certain axioms [23]: any function that fails to satisfy these axioms is *not* a measure of information. Many of these properties are "obviously" satisfied by the form in Eq. (1). This form, however, might not satisfy some of the other defining properties of information, namely: [23,24]

(a) Nonnegativity The change in information is nonnegative,

$$I\left[p(\boldsymbol{x})|\,p^0(\boldsymbol{x})\right] \equiv \int p(\boldsymbol{x}) \ln\left(\tfrac{p(\boldsymbol{x})}{p^0(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \geq 0 \tag{3}$$

and is zero only if the reference distribution and the final distribution are identical.

(b) Additivity If two variables are independent, then the total change in Information is the sum of the changes in information of the separate variables.

$I\left[p|\,p^0\right]$ is called the Kullback–Liebler information, and measures the information loss (or "missing information") in $p(\boldsymbol{x})$ relative to a prior distribution, $p^0(\boldsymbol{x})$. Equivalently, $I\left[p|\,p^0\right]$ denotes the entropy gain (relative to the prior distribution) or the conditional information: the amount of information contained in $p(\boldsymbol{x})$, given that $p^0(\boldsymbol{x})$ was already known.

The derivations of these results is instructive. To prove nonnegativity, one uses the inequality [25]

$$\ln(K) \geq 1 - \tfrac{1}{K} \tag{4}$$

to deduce that, for the Kullback–Liebler conditional information,

$$\begin{aligned} \int p(\boldsymbol{x}) \ln\left(\tfrac{p(\boldsymbol{x})}{p^0(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} &\geq \int p(\boldsymbol{x})\left(1 - \tfrac{p^0(\boldsymbol{x})}{p(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \\ &\geq \int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \int p^0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &\geq 0 \end{aligned} \tag{5}$$

Note that the last step is valid only if

$$\int p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int p^0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{6}$$

and that, because of the possibility of interatomic electron transfer, Eq. (6) does not generally hold for the electron densities used in Eq. (1). For example, if atom B is less electronegative than other atoms in the molecule, then charge transfer from atom B to the other atoms in the molecule gives

$$N_B \equiv \int \rho_B(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} < \int \rho_B^0(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \equiv N_B^0, \tag{7}$$

which violates Eq. (6).

It is not at all clear, then, that the information as defined in Eq. (1) is always positive. An electron donor's density may be expected to decrease almost everywhere during the course of molecule formation, giving $\ln\left(\frac{\rho_B(\boldsymbol{r})}{\rho_B^0(\boldsymbol{r})}\right) < 0$ for most points, $\boldsymbol{r}$. This, paradoxically, predicts that atom B *gains* information during molecule formation. (One would expect that, in the process of "perturbing" an atom from its unbound state, the missing information would increase, not decrease.) Just because one or more of the terms in the summand in Eq. (1) is negative, however, does not mean that the entire summand is negative. However, before Eq. (1) can be established as a measure of information, it must be *proved* that

$$I\left[\{\rho_A(\boldsymbol{r})\}_{A=1}^M \middle| \left\{\rho_A^0(\boldsymbol{r})\right\}_{A=1}^M\right] \geq 0. \tag{8}$$

If Eq. (8) is not satisfied, the Hirshfeld atom is not truly information-theoretic (though one might argue that it is "in the spirit of information theory").

The fact that the entropy due to independent events must be the sum of the entropies of the separate events is intuitively obvious and, in a continuous context, is manifest when a probability distribution function can be written as a product of two pieces, $p(\boldsymbol{x}) = f(\phi)g(\gamma)$. One then has that

$$\begin{aligned} \int p(\boldsymbol{x}) \ln\left(\tfrac{p(\boldsymbol{x})}{p^0(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} &= \int f(\phi)g(\gamma) \ln\left(\frac{f(\phi)g(\gamma)}{f^0(\phi)g^0(\gamma)}\right) \mathrm{d}\phi \mathrm{d}\gamma \\ &= \int f(\phi)g(\gamma) \ln\left(\frac{f(\phi)}{f^0(\phi)}\right) \mathrm{d}\phi \mathrm{d}\gamma \\ &\quad + \int f(\phi)g(\gamma) \ln\left(\frac{g(\gamma)}{g^0(\gamma)}\right) \mathrm{d}\phi \mathrm{d}\gamma \\ &= \int g(\gamma) \mathrm{d}\gamma \int f(\phi) \ln\left(\frac{f(\phi)}{f^0(\phi)}\right) \mathrm{d}\phi \\ &\quad + \int f(\phi) \mathrm{d}\phi \int g(\gamma) \ln\left(\frac{g(\gamma)}{g^0(\gamma)}\right) \mathrm{d}\gamma \\ &= \int f(\phi) \ln\left(\frac{f(\phi)}{f^0(\phi)}\right) \mathrm{d}\phi \\ &\quad + \int g(\gamma) \ln\left(\frac{g(\gamma)}{g^0(\gamma)}\right) \mathrm{d}\gamma. \end{aligned} \tag{9}$$

The last step in this derivation is only possible because, since $f(\phi)$ is the probability of observing *some* value of the variable ϕ, it must be that

$$\int f(\phi) \mathrm{d}\phi = 1 = \int g(\gamma) \mathrm{d}\gamma. \tag{10}$$

It is again unclear as to whether the additivity property of information is satisfied by the measure of information in Eq.

(1). One might suspect this is a purely academic point—that in "real" molecules one can never write the density as a product of two factors, but this is not the case. Due to symmetry, in a diatomic molecule, we can write the atomic densities as a product of two functions,

$$\rho_A(\boldsymbol{r}) = \sigma_A(r_A, \theta_A) f_A(\phi_A). \tag{11}$$

Here, $\sigma_A(r_A, \theta_A)$ depends on the distance from the nuclear centre, r_A, and the angle in the plane of the internuclear axis, θ_A, and $f_A(\phi_A)$ depends on the angle about the internuclear axis, ϕ_A. $f_A(\phi_A)$ is often simply $\frac{1}{2\pi}$, but is not constant for molecular states with nonzero total orbital angular momentum quantum numbers. According to the independence principle, the fact that electronic motion about the internuclear axis is independent of their motion in other directions[1] should allow us to write the information as a sum of two terms of these types but, because the electron densities in Eq. (1) is not normalized to unity, the derivation fails.

2.2 Alternative measures of information

There are a number of different definitions for the information that are available, including several that do not satisfy the axiom of independence. In physics parlance, such entropies are said to be nonextensive: they do not grow linearly as the number of "independent" systems increases. (By contrast, the thermodynamic entropy from classical thermodynamics is clearly extensive.) One such entropy is the Tsallis, or Hartley, entropy: [26]

$$S_T[p(\boldsymbol{x})] = \int p(\boldsymbol{x}) \frac{(p(\boldsymbol{x}))^{q-1} - 1}{q-1} \mathrm{d}\boldsymbol{x}. \tag{12}$$

Equation (12) reduces to the Shannon form of the entropy, $S_S[p(\boldsymbol{x})] = -\int p(\boldsymbol{x}) \ln p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ in the limit as q approaches 1. Except in this limit, the axiom of independence does not hold. (This is apparent from the derivation in Eq. (9), which hinges on the fact that the logarithm of a product is the sum of the logarithms.)

Motivated by the Tsallis form of the entropy [27–31], one can measure the information distance between two probability distribution functions using the analogy to the Kullback–Liebler form, so that

$$I_T\left[p(\boldsymbol{x}) \middle| p^0(\boldsymbol{x})\right] = \int p(\boldsymbol{x}) \frac{\left(\frac{p(\boldsymbol{x})}{p^0(\boldsymbol{x})}\right)^{q-1} - 1}{q-1} \mathrm{d}\boldsymbol{x}. \tag{13}$$

The author knows of no proof for the nonnegativity of Eq. (13), which is probably one reason why this is not the preferred definition of the Kullback–Liebler conditional information in Tsallis statistics, namely [29,30]

$$I_T^{\text{preferred}}\left[p(\boldsymbol{x}) \middle| p^0(\boldsymbol{x})\right] = \frac{\int p(\boldsymbol{x}) \left[\frac{(p(\boldsymbol{x}))^{q-1}-1}{q-1} - \frac{(p^0(\boldsymbol{x}))^{q-1}-1}{q-1}\right] \mathrm{d}\boldsymbol{x}}{\int (p(\boldsymbol{x}))^q \mathrm{d}\boldsymbol{x}} \tag{14}$$

Unlike Eq. (14), however, substituting Eq. (13) for the Kullback–Liebler information in Eq. (1) leads to the Hirshfeld atom. Minimizing

$$I_T\left[\{\rho_A(\boldsymbol{r})\}_{A=1}^M \middle| \{\rho_A^0(\boldsymbol{r})\}_{A=1}^M\right] = \sum_{A=1}^M \int \rho_A(\boldsymbol{r}) \frac{\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)^{q-1} - 1}{q-1} \mathrm{d}\boldsymbol{r} \tag{15}$$

using a Lagrange multiplier, $\lambda(\boldsymbol{r})$, to force the Eq. (2), one has the stationary conditions

$$0 = \frac{\delta\left(I_T\left[\{\rho_A(\boldsymbol{r})\}_{A=1}^M \middle| \{\rho_A^0(\boldsymbol{r})\}_{A=1}^M\right] - \lambda(\boldsymbol{r})\left[\sum_{A=1}^M \rho_A(\boldsymbol{r}) - \rho_{\text{mol}}(\boldsymbol{r})\right]\right)}{\delta\rho_A(\boldsymbol{r})}$$

$$\lambda(\boldsymbol{r}) = \left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)^{q-1} \left(\frac{1}{q-1} + 1\right) \tag{16}$$

which leads to the conclusion that

$$\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})} = \frac{\rho_B(\boldsymbol{r})}{\rho_B^0(\boldsymbol{r})} \qquad 1 \leq A, B \leq M. \tag{17}$$

Elementary manipulation using Eq. (2) [cf. Eqs. (42)–(44) in [32]] leads to the Hirshfeld result,

$$\rho_A(\boldsymbol{r}) = \frac{\rho_A^0}{\sum\limits_{B=1}^M \rho_B^0(\boldsymbol{r})} \rho_{\text{mol}}(\boldsymbol{r}). \tag{18}$$

While the preceding analysis may be taken as an argument for the universality of the Hirshfeld form, it should not be assumed that every reasonable choice for the Kullback–Liebler conditional information leads to the Hirshfeld atom. This is manifestly not the case for the Rajagopal–Abe form [Eq. (14)] and, as shown in the Appendix, the Fisher information measure does not lead to the Hirshfeld atom either.

3 Entropy of mixing and the Hirshfeld atom

The principal weakness in the traditional "information-theoretic" of the Hirshfeld atom is their nebulous relationship to bona fide information theory. In particular, it has not been demonstrated that the information measure in Eq. (1),

$$I\left[\{\rho_A(\boldsymbol{r})\}_{A=1}^M \middle| \{\rho_A^0(\boldsymbol{r})\}_{A=1}^M\right] = \sum_{A=1}^M \int \rho_A(\boldsymbol{r}) \ln \frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})} \mathrm{d}\boldsymbol{r} \tag{19}$$

satisfies the defining properties of the information, notably the requirements of nonnegativity and the additivity of the

---

[1] This, of course, is not rigorously true for many-electron molecules, where this symmetry property of the electron density would not be observed in the electron pair density. (The electron pair density has symmetries of its own, though.) But for a one electron molecule, this is certainly the case

information from independent random variables. The question arises: is Eq. (19) a measure of the information loss upon molecule formation or is it merely an "information-theory motivated" measure of the changes associated with molecule formation? That is, can Eq. (19) be placed on a firm footing from the standpoint of information theory, or is it just an information-theoretic heuristic?

From the analysis in Sect. 2.1, the derivation of many key properties of the Kullback–Liebler relative information relies on the fact that statistical probability distribution functions are normalized to unity. This suggests that instead of considering comparing the electron densities of the atom in the molecule to a reference atom, we compare their shape functions. The shape function is the density per particle [33,34],

$$\sigma(\boldsymbol{r}) = \frac{\rho(\boldsymbol{r})}{\int \rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r}}, \tag{20}$$

and, combined with the number of electrons,

$$N[\rho] \equiv \int \rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{21}$$

fully specifies the state of any system. We propose, then, that the information loss during molecule formation can be modeled as

$$I\Big[\{p_A,\sigma_A\}_{A=1}^M \Big| \{p_A^0,\sigma_A^0\}_{A=1}^M\Big] = \sum_{A=1}^M p_A \int \sigma_A(\boldsymbol{r}) \ln\left(\frac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r} + \sum_{A=1}^M p_A \ln\left(\frac{p_A}{p_A^0}\right), \tag{22}$$

where N_{mol} is the number of electrons in the molecule, $p_A = \frac{N_A}{N_{\text{mol}}}$ is the proportion of the electrons in the molecule that are assigned to a given atom (N_A is the number of electrons in atom A) and $\sigma_A(r)$ is the shape function of the atom in the molecule. For the reference systems, $\sigma_A^0(r)$ and p_A^0 denote the shape function and the fraction of the system's electrons contained in the reference atom, respectively.

Exploring Eq. (22) term by term, the first term,

$$I_{\text{shape}}\big[\sigma_A \,|\, \sigma_A^0\big] \equiv p_A \int \sigma_A(\boldsymbol{r}) \ln\left(\frac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r}, \tag{23}$$

clearly measures the information lost due to the change in shape due to molecule formation. The factor of p_A can be justified on intuitive grounds: it is reasonable to suspect that the "total information loss" due to the shape deformation is proportional to the fraction of electrons that are contained in that atom, so that distortions in the shape function of a "heavy" atom will be proportionately more important than that with fewer electrons. Another, equally important, rationale is that this method for weighting the change in the shape function gives the first and second term in Eq. (22) a similar form, which allows the coefficient of p_A to be interpreted as the change in information of atom A during molecule formation. There is also a rigorous mathematical justification for the p_A coefficient. Duplicates of a system impart no additional information, because all the information that is contained in the "copies" was already apparent from the original and, were the factor of p_A omitted in Eq. (23), this result would no longer be obtained. Note that because the shape function is normalized to unity, $I\big[\sigma_A \,|\, \sigma_A^0\big]$ is never negative. In addition, the additivity of the information in independent events is preserved.

The second term in Eq. (22) is an "entropy of mixing" term,

$$I_{\text{mix}}\Big[\{p_A\}_{A=1}^M \Big| \{p_A^0\}_{A=1}^M\Big] \equiv \sum_{A=1}^M p_A \ln\left(\frac{p_A}{p_A^0}\right) \tag{24}$$

and is directly analogous in form and interpretation to the eponymous concept in the classical thermodynamics of mixtures. The entropy of mixing term is essential: the possibility of charge transfer between atoms dictates that just because the shape of an atom does not change during molecule formation, it does not mean that the atom does not change, and so information can be lost during molecule formation even if the shapes of all the molecule's composing atoms are unchanged, $\sigma_A(\boldsymbol{r}) = \sigma_A^0(\boldsymbol{r})$. Because

$$1 = \sum_{A=1}^M p_A = \sum_{A=1}^M p_A^0 \tag{25}$$

the "mixing" information is always positive and satisfies the requirement that independent events lead to additive information. (These results are the discrete analogue to statements (a) and (b) in Sect. 2.1. To prove them, merely replace the integrals in Eqs. (5) and (9) with summations.)

When considering the entropy gain upon molecule formation instead of the information lost, a slight revision to Eq. (22) should be made. The entropy is an extensive quantity and, in contrast to the information, increases in proportion to the number of distinct "copies" of a molecule under consideration system. Multiplication of the sum of the information lost by the total number of electrons in the system, N_{mol}, gives a measure of the entropy gain during molecule formation [35],

$$\begin{aligned} &S\Big[\{p_A,\sigma_A\}_{A=1}^M \Big| \{p_A^0,\sigma_A^0\}_{A=1}^M\Big] \\ &\quad\equiv N_{\text{mol}} \cdot I\Big[\{p_A,\sigma_A\}_{A=1}^M \Big| \{p_A^0,\sigma_A^0\}_{A=1}^M\Big] \\ &\quad= \sum_{A=1}^M N_A \int \sigma_A(\boldsymbol{r}) \ln\left(\frac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r} \\ &\qquad+ \sum_{A=1}^M N_A \ln\left(\frac{p_A}{p_A^0}\right). \end{aligned} \tag{26}$$

Multiplication of the information by a constant clearly does not change the particular information-theoretic atoms so derived, as it amounts to nothing more than a change in units. Using Eq. (26), if we consider $\mathcal{K}$ well-separated copies of a molecule, the "entropy of molecule formation" for the entire system will be $\mathcal{K}$ times that of a single molecule, in accordance with our chemical intuition. Without the factor of N_{mol}, one finds that the information lost during molecule formation is independent of the number of copies of the molecule.

Expression (26) for the entropy of molecular formation relative to some reference is easily constructed from its components. The key assumption is that the shape functions, $\{\sigma_A(\boldsymbol{r})\}$, and electron fractions, $\{p_A\}$, of the atoms can be varied independently. Then, according to the postulate that the entropy of independent variables is additive, Eq. (26) results. From Eq. (26) and the definition of the shape function, Eq. (20),

$$\begin{aligned} &S\Big[\{p_A,\sigma_A\}_{A=1}^{M}\Big|\{p_A^0,\sigma_A^0\}_{A=1}^{M}\Big] \\ &= N_{\mathrm{mol}}\left(\sum_{A=1}^{M} p_A \int \sigma_A(\boldsymbol{r}) \ln\left(\tfrac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r} + \sum_{A=1}^{M} p_A \ln\left(\tfrac{p_A}{p_A^0}\right)\right) \\ &= N_{\mathrm{mol}}\left(\sum_{A=1}^{M} p_A \left[\int \sigma_A(\boldsymbol{r}) \ln\left(\tfrac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r} + \int \sigma_A(\boldsymbol{r}) \ln\left(\tfrac{p_A}{p_A^0}\right)\mathrm{d}\boldsymbol{r}\right]\right) \\ &= \left(\sum_{A=1}^{M} N_{\mathrm{mol}}\left(\tfrac{N_A}{N_{\mathrm{mol}}}\right)\left[\int \sigma_A(\boldsymbol{r}) \ln\left(\tfrac{p_A\sigma_A(\boldsymbol{r})}{p_A^0\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r}\right]\right) \\ &= \left(\sum_{A=1}^{M} N_A\left[\int \sigma_A(\boldsymbol{r})\left(\ln\left(\tfrac{N_A\sigma_A(\boldsymbol{r})}{N_A^0\sigma_A^0(\boldsymbol{r})}\right) + \ln\left(\tfrac{N_{\mathrm{mol}}^0}{N_{\mathrm{mol}}}\right)\right)\mathrm{d}\boldsymbol{r}\right]\right) \\ &= -N_{\mathrm{mol}} \ln \tfrac{N_{\mathrm{mol}}}{N_{\mathrm{mol}}^0} + \sum_{A=1}^{M} \int \rho_A(\boldsymbol{r}) \ln\left(\tfrac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r}. \end{aligned} \tag{27}$$

The first term in Eq. (27) is constant, and is usually zero (unless sums of the charges of the reference atoms do not add up to the total molecular charge). The second term is exactly the Nalewajski–Parr expression for the entropy, Eq. (1) [22]. Minimizing Eq. (27) subject to the isodensity constraint, Eq. (2), then yields the Hirshfeld partitioning, Eq. (18).

The importance of Eq. (27) is clear. We have started with a mathematical measure of the information lost during molecule formation and from it derived the Nalewajski–Parr entropy, which was originally proposed based on more heuristic considerations. This reveals, in particular, that the Nalewajski–Parr entropy is a "real" information-theoretic measure, never negative and preserving the additive entropies of independent events. That is, the Hirshfeld atom is really an "information-theoretic" atom in a rigorous mathematical sense. Another advantage of Eq. (26) over the original formulation [Eq. (1) or (27)] is that the loss of information due to charge transfer or "electron mixing," Eq. (24), has been separated from the loss of information due to polarization (an atom's shape), Eq. (23). This is expected to have conceptual utility [35].

The primary motivation for presenting the Tsallis form of the conditional entropy, Eq. (13), is to demonstrate that the preceding result is not a trivial one. The derivation in Eq. (27) uses the fact that the logarithm of a product is the sum of the logarithms, and so this derivation does not work for the nonlogarithmic Tsallis form [cf. Eq. (13)] – a rigorous information-theoretic treatment based on the Tsallis form leads to a different definition of the atom in a molecule, and not the Hirshfeld form. (A closed form solution, similar to Eq. (18), does not seem to be attainable in this case.) Nor does the present method of derivation work for the Rajagopal–Abe form [Eq. (14)] or the Fisher information (see the Appendix).

Concluding this section, it is clear from both the Eqs. (26) and (1) that the entropy gained during molecule formation can be associated with the quantity

$$\begin{aligned} &S_A\big[p_A,\sigma_A\big|\,p_A^0,\sigma_A^0\big] \\ &= N_{\mathrm{mol}} p_A\left(\int \sigma_A(\boldsymbol{r}) \ln\left(\tfrac{\sigma_A(\boldsymbol{r})}{\sigma_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r} + \ln\left(\tfrac{p_A}{p_A^0}\right)\right) \\ &= \int \rho_A(\boldsymbol{r}) \ln\left(\tfrac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\mathrm{d}\boldsymbol{r}. \end{aligned} \tag{28}$$

It is worth noting that, as per the discussion in Sect. 2.1, $S_A\big[p_A,\sigma_A\big|\,p_A^0,\sigma_A^0\big]$ can actually be less than zero. Given that the shape-function entropy is never negative, this has a clear interpretation: an atom that loses electrons upon molecule formation can "lose" information. This coincides with the observation, first noted by Sears et al. [36], that electronic motion is intertwined with entropy. Thus, an atom that loses electrons can also "gain" information (or lose uncertainty).

4 Atoms from local conditional probabilities

Nalewajski has asserted that the Hirshfeld partitioning is entirely general, and can be derived from many different measures of information. To arrive at this conclusion, in much of his recent work an alternative approach to the Hirshfeld partitioning has been used [13–15,21]. Indeed, with this alternative method, one can derive the Hirshfeld partitioning from *any* information measure, as we now rigorously show.

The alternative approach advocated by Nalewajski is based on the probability that an electron at the point $\boldsymbol{x}$ is assigned to atom A,

$$\mathcal{P}_A(\boldsymbol{x}) \equiv \frac{\rho_A(\boldsymbol{x})}{\rho_{\mathrm{mol}}(\boldsymbol{x})} \tag{29}$$

Here, as before, $\rho_A(\boldsymbol{x})$ denotes the electron density of atom A and $\rho_{\mathrm{mol}}(\boldsymbol{x})$ denotes the electron density of the molecule as a whole. The fact that the atomic densities must add up to form the molecular density, Eq. (2), imposes the normalization constraint

$$\sum_{A=1}^{M} \mathcal{P}_A(\boldsymbol{x}) = 1, \tag{30}$$

while the fact that both atomic densities and the total molecular density are nonnegative implies that $\mathcal{P}_A(\boldsymbol{x}) \geq 0$. Based on these properties, we conclude that a vector containing the value of $\mathcal{P}_A(\boldsymbol{x})$ for each A,

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) \equiv [\mathcal{P}_1(\boldsymbol{x}), \mathcal{P}_2(\boldsymbol{x}), \ldots, \mathcal{P}_M(\boldsymbol{x})]^T, \tag{31}$$

is a valid discrete probability distribution function. We denote the prior distribution in a similar way, with $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x}) \equiv \big[\mathcal{P}_1^0(\boldsymbol{x}), \mathcal{P}_2^0(\boldsymbol{x}), \ldots, \mathcal{P}_M^0(\boldsymbol{x})\big]^T$.

In general, a measure of information satisfies several properties, each of which is chosen to be intuitively reasonable and useful in a given situation. On intuitive grounds, however, *every* measure of the missing information is a directed divergence [23]. That is, the quantity of "missing information," $\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{P}}^0\right]$, must be positive whenever $\boldsymbol{\mathcal{P}} \neq \boldsymbol{\mathcal{P}}^0$, and zero when $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{P}}^0$. Any function,

$$\begin{aligned} &\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{P}}^0\right] = 0 \quad \boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{P}}^0 \\ &\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{P}}^0\right] > 0 \quad \boldsymbol{\mathcal{P}} \neq \boldsymbol{\mathcal{P}}^0 \end{aligned} \tag{32}$$

that satisfies these constraints is a directed divergence. Note, however, that while every measure of information is a directed divergence, it is unreasonable to assume that every directed divergence is a measure of information. (The conditions in Eq. (32) are not very restrictive, so some functions that are directed divergences are too ill-behaved to be reasonable information measures. In addition, every measure of distance is necessarily a directed divergence, so any measure of distance will satisfy Eq. (32).) All of the measures of information considered previously also satisfy Eq. (32). Of special interest is the Kullback–Liebler information [cf. Eq. (24)]

$$I\left[\boldsymbol{\mathcal{P}}\,|\boldsymbol{\mathcal{P}}^0\right] \equiv \sum_{A=1}^{M} \mathcal{P}_A(\boldsymbol{x}) \ln\left(\frac{\mathcal{P}_A(\boldsymbol{x})}{\mathcal{P}_A^0(\boldsymbol{x})}\right), \tag{33}$$

which has been the subject of much prior work [13–15,21]. Because the distributions under consideration are appropriately normalized [cf. Eq. (30)], Eq. (33) is a mathematically acceptable measure of information.

In order to find the atom that minimizes the information measure (more generally, the directed divergence), we must minimize $\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{P}}^0\right]$ subject to the constraint that $\mathcal{P}(\boldsymbol{x})$ is a valid probability distribution function. Specifically,

$$\mathcal{I}_{\min}(\boldsymbol{x}) \equiv \underbrace{\min}_{\substack{0\le\mathcal{P}_A(\boldsymbol{x}) \\ 1=\sum\limits_{A=1}^{M}\mathcal{P}_A(\boldsymbol{x})}} \quad \mathcal{I}\left[\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{P}}^0\right] \tag{34}$$

From Eq. (32), it is clear that $\mathcal{I}_{\min}(\boldsymbol{x}) \geq 0$, and that $\mathcal{I}_{\min}(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$. However, since $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$ is a valid probability distribution function, choosing $\mathcal{P}_A(\boldsymbol{x}) = \mathcal{P}_A^0(\boldsymbol{x})$ for all A satisfies the constraints on the minimization. Thus, the constaints do not "restrict" the minimization, so $\mathcal{I}_{\min}(\boldsymbol{x}) = 0$ and $\mathcal{P}_A(\boldsymbol{x}) = \mathcal{P}_A^0(\boldsymbol{x})$ for all $\boldsymbol{x}$. This should be contrasted with the information measure originally used by Nalewajski and Parr (and rederived in Sect. 3.); when minimizing Eq. (1) [Eq. (26)] the requirement that the sum of the atomic densities equals the molecular density [Eq. (2)] generally prevents the missing information, $I\left[\{\rho_A(\boldsymbol{r})\}_{A=1}^{M}\middle|\{\rho_A^0(\boldsymbol{r})\}_{A=1}^{M}\right]$, from being zero.

If one chooses the prior distribution according to Hirshfeld's stockholder principle, whereby the probability of an electron being assigned to a given atom is proportional to that atom's "investment" to the density of the pro-molecule,

$$\mathcal{P}_A^0(\boldsymbol{x}) \equiv \frac{\rho_A^0(\boldsymbol{x})}{\sum\limits_{B=1}^{M}\rho_B^0(\boldsymbol{x})}, \tag{35}$$

then, by the preceding argument, any measure of information, $\mathcal{I}\left[\mathcal{P}|\,\mathcal{P}^0\right]$, yields the Hirshfeld partitioning. Specifically, from Eq. (29),

$$\begin{aligned} \mathcal{P}_A(\boldsymbol{x}) &= \frac{\rho_A(\boldsymbol{x})}{\rho_{\text{mol}}(\boldsymbol{x})} = \frac{\rho_A^0(\boldsymbol{x})}{\sum\limits_{B=1}^{M}\rho_B^0(\boldsymbol{x})} = \mathcal{P}_A^0(\boldsymbol{x}) \\ \rho_A(\boldsymbol{x}) &= \frac{\rho_A^0(\boldsymbol{x})\rho_{\text{mol}}(\boldsymbol{x})}{\sum\limits_{B=1}^{M}\rho_B^0(\boldsymbol{x})}. \end{aligned} \tag{36}$$

Equation (36) is not a trivial result; the fact that the "output" distribution from the procedure, $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$, is equal to the "input" distribution, $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$, indicates that the input distribution is "stable" with respect to the minimization in Eq. (34). It should be noted that this stability does not depend on the specific information measure under consideration. Neither does this stability depend on the prior distribution, $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$ – any prior distribution that is a bona fide probability distribution function (that is, $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$ must be nonnegative and normalized) will work. On the basis of the preceding analysis, we can make the general statement: *For any directed divergence,* $\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\,\boldsymbol{\mathcal{P}}^0\right]$*, and prior distribution,* $\boldsymbol{\mathcal{P}}^0(\boldsymbol{x})$*, minimizing of* $\mathcal{I}\left[\boldsymbol{\mathcal{P}}|\,\boldsymbol{\mathcal{P}}^0\right]$ *with respect to the constraint that the function* $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ *be nonnegative and normalized yields the atomic density,*

$$\rho_A(\boldsymbol{x}) = \mathcal{P}_A^0(\boldsymbol{x})\rho_{\text{mol}}(\boldsymbol{x}) \tag{37}$$

The Hirshfeld atom is derived when $\mathcal{P}^0(\boldsymbol{x})$ is defined with Eq. (35). Taking a different prior distribution, e.g.,

$$\mathcal{P}_A^0(\boldsymbol{x}) \equiv \frac{Z_A|\boldsymbol{x}-\boldsymbol{R}_A|^2}{\sum\limits_{B=1}^{M} Z_B|\boldsymbol{x}-\boldsymbol{R}_B|^2}, \tag{38}$$

yields an atom based on the relative amount of force exerted on an electron at $\boldsymbol{x}$ from the atomic nuclei (with atomic numbers $\{Z_A\}$ and positions $\{\boldsymbol{R}_A\}$),

$$\rho_A^{\text{force}} \equiv \rho_{\text{mol}}(\boldsymbol{x})\frac{Z_A|\boldsymbol{x}-\boldsymbol{R}_A|^2}{\sum\limits_{B=1}^{M} Z_B|\boldsymbol{x}-\boldsymbol{R}_B|^2}. \tag{39}$$

Alternatively, if we choose

$$\mathcal{P}_A^0(\boldsymbol{x}) \equiv \begin{cases} 1 & \boldsymbol{x} \in \Omega_A, \\ 0 & \boldsymbol{x} \notin \Omega_A, \end{cases} \tag{40}$$

where Ω_A is an atomic volume derived using Bader's method [37,38], then one obtains the Bader's topological partitioning.

The method sketched in Eqs. (29)–(37), which we term the local conditional distribution method, can be contrasted with the method used in Sect. 3, wherein integration "adds up" the contributions to the entropy from all points in space to provide a "global" measure of the entropy. Also unlike the local partitioning scheme, the "entropy of mixing" argument in section 3, is specifically associated with the Hirshfeld atom: non-logarithmic measures of the information – to say nothing of other measures of the directed divergence – are associated with different definitions of the atom. In addition, and again unlike the local conditional distribution method [cf. Eq. (35)], the stockholder analysis does not directly enter into the derivation of "entropy of mixing" argument. Thus, based on the argument in Sect. 3, then, we identify the Hirshfeld atom as *the* information-theoretic atom or, rather, *the Kullback–Liebler information-theoretic* atom. The elucidation of this relationship is the primary contribution of this paper.

5 Summary

This paper addresses, in a rigorous mathematical way, the uncertainties associated with the information-theoretic approach to the Hirshfeld atom. By recasting the original formulation of Nalewajski and Parr [22] in terms of the shape function and an "entropy of mixing" term, one can rigorously derive the Hirshfeld partitioning. [1] Importantly, this demonstrates that the Nalewajski–Parr information measure satisfies important constraints associated with nonnegativity and statistical independence. It also seems likely that decomposing the relative information into polarization (shape deformation) and charge transfer (mixing) components may be useful. In particular, we learn that atoms which lose electrons during molecule formation will often, somewhat counterintuitively, "gain information" relative to their free state.

While the Hirshfeld atom seems to be the "best" atom from the standpoint of Shannon's information theory, the same cannot be said when alternative measures of information are used. In particular, even though a naïve derivation using the Tsallis information will give the Hirshfeld atom, this result disintegrates when a more rigorous mathematical track is taken. The same may be said of the Fischer information and, presumably, many other methods of measuring the information also. This, however, does not diminish the importance of Hirshfeld atom or its information-theoretic derivation. Rather, it merely serves to emphasize the obvious: there are many different ways to define an atom in a molecule and, depending on the criteria used, different definitions result. This is beneficial: one may choose, in any given situation, the definition for an atom in a molecule that is most convenient, useful, and appropriate. The Hirshfeld atom and the entropy of mixing argument presented in Sect. 3 are simple and intuitive, however, and might plausibly be preferred to other, more complicated, alternatives.

6 Appendix

Nalewajski and Parr have asserted that the relative Fisher information,

$$I^F\left[\{\rho_A\}_{A=1}^M \left| \{\rho_A^0\}_{A=1}^M \right.\right] = \sum_{A=1}^{M} \left\langle \rho_A(\boldsymbol{r}) \left| \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right|^2 \right\rangle, \tag{41}$$

can, like the Shannon information, be used to derive the Hirshfeld atom [14,17]. Taking the functional derivative of expressions like Eq. (41) is, however, quite difficult, and subject to error. In particular, before using formula like those of Gelfand and Fomin [39] one must be careful to ensure that the assumptions implicit in the derivation of those formulas are satisfied. The surest way to proceed is to use the definition of the functional derivative, by which we identify $\frac{\delta I^F}{\delta \rho_A(\boldsymbol{r})}$ with the coefficient of $\delta\rho_A(\boldsymbol{r})$ in the Taylor series expansion of $I^F\left[\{\rho_A + \delta\rho_A(\boldsymbol{r})\}_{A=1}^M \left| \{\rho_A^0\}_{A=1}^M \right.\right] - I^F\left[\{\rho_A\}_{A=1}^M \left| \{\rho_A^0\}_{A=1}^M \right.\right]$. Proceeding in this manner, and neglecting terms of higher-order in $\delta\rho(\boldsymbol{r})$ whenever they appear, we have

$$\begin{aligned}
& I^F\left[\{\rho_A + \delta\rho_A(\boldsymbol{r})\}_{A=1}^M \left| \{\rho_A^0\}_{A=1}^M \right.\right] - I^F\left[\{\rho_A\}_{A=1}^M \left| \{\rho_A^0\}_{A=1}^M \right.\right] \\
&= \left\langle (\rho_A(\boldsymbol{r}) + \delta\rho(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})+\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})+\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad - \left\langle (\rho_A(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&= \left\langle (\rho_A(\boldsymbol{r}) + \delta\rho(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\left(1 + \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right)\right]\right. \\
&\quad \left. \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\left(1 + \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right)\right]\right\rangle \\
&\quad - \left\langle (\rho_A(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&= \left\langle (\rho_A(\boldsymbol{r}) + \delta\rho(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right) + \ln\left(1 + \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right. \\
&\quad \left. \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right) + \ln\left(1 + \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad - \left\langle (\rho_A(\boldsymbol{r}))\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&= \left\langle \delta\rho(\boldsymbol{r})\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad + 2\left\langle \rho_A(\boldsymbol{r})\nabla\left[\ln\left(1 + \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&= \left\langle \delta\rho(\boldsymbol{r})\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad + 2\left\langle \rho_A(\boldsymbol{r})\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right]\right\rangle.
\end{aligned} \tag{42}$$

The last line uses the Taylor series for the logarithm, $\ln(1+x) = x - \frac{x^2}{2} + \ldots$. Next, from the identity $\nabla^2(f(\boldsymbol{x})g(\boldsymbol{x})) = g(\boldsymbol{x})\nabla^2 f(\boldsymbol{x}) + 2\nabla g(\boldsymbol{x}) \cdot \nabla f(\boldsymbol{x}) + f(\boldsymbol{x})\nabla^2 g(\boldsymbol{x})$, one obtains

$$\begin{aligned}
& 2\left\langle \rho_A(\boldsymbol{r})\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \cdot \nabla\left[\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right]\right\rangle \\
&= \left\langle \rho_A(\boldsymbol{r})\nabla^2\left[\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad - \left\langle \rho_A(\boldsymbol{r})\left(\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla^2\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad - \left\langle \rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla^2\left(\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right\rangle
\end{aligned} \tag{43}$$

and, applying Green's theorem, one obtains

$$\begin{aligned}
&2\left\langle \rho_A(\boldsymbol{r})\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\cdot\nabla\left[\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right]\right\rangle \\
&= \left\langle \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla^2\rho_A(\boldsymbol{r})\right\rangle \\
&\quad + \oiint_{R\to\infty}\begin{pmatrix} \rho_A(\boldsymbol{r})\nabla\left(\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right) \\ -\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla\rho_A(\boldsymbol{r}) \end{pmatrix}\cdot\boldsymbol{n}\,\mathrm{d}S \\
&\quad -\left\langle \rho_A(\boldsymbol{r})\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\nabla^2\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad -\left\langle \frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\nabla^2\left[\rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right\rangle \\
&\quad - \oiint_{R\to\infty}\begin{pmatrix} -\rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla\left(\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right) \\ -\frac{\delta\rho(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\nabla\left(\rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right) \end{pmatrix}\cdot\boldsymbol{n}\,\mathrm{d}S \\
&= \left\langle \frac{\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla^2\rho_A(\boldsymbol{r}) - \rho_A(\boldsymbol{r})\nabla^2\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] - \nabla^2\left[\rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]}{\rho_A^0(\boldsymbol{r})}\delta\rho(\boldsymbol{r})\right\rangle \\
&\quad +2\oiint_{R\to\infty}\left(\delta\rho(\boldsymbol{r})\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\right)\cdot\boldsymbol{n}\,\mathrm{d}S, \qquad (44)
\end{aligned}$$

where it is understood that the surface integral occurs over a surface located at infinity. For highly local perturbations,[2] the surface integral will be zero. Using Eq. (44) in Eq. (42) allows one to identify the functional derivative, namely,

$$\begin{aligned}
&\frac{\delta I_F}{\delta\rho_A(\boldsymbol{r})} \\
&= \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\cdot\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \\
&\quad + \frac{\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\nabla^2\rho_A(\boldsymbol{r}) - \rho_A(\boldsymbol{r})\nabla^2\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] - \nabla^2\left[\rho_A(\boldsymbol{r})\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]}{\rho_A^0(\boldsymbol{r})} \\
&= \nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right]\cdot\nabla\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] - 2\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\nabla^2\left[\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right)\right] \\
&\quad -2\frac{\nabla\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\cdot\nabla\ln\left(\frac{\rho_A(\boldsymbol{r})}{\rho_A^0(\boldsymbol{r})}\right) \qquad (45)
\end{aligned}$$

Minimizing the Fisher relative information functional, Eq. (41) subject to the molecular density constraint, Eq. (2), yields the identity

$$\frac{\delta I_F}{\delta\rho_A(\boldsymbol{r})} = \frac{\delta I_F}{\delta\rho_B(\boldsymbol{r})} \qquad 1 \le A, B \le M. \qquad (46)$$

Perusal of Eq. (45) reveals that, because of the form of the last term, we do not obtain Eq. (17), which is the key result required to derive the Hirshfeld atom. Consequently, the minimum Fisher information principle does not lead to the Hirshfeld atom. Indeed, it seems impossible to solve Eq. (46) to get any elementary formula for the atomic definition corresponding to the minimum Fisher information principle. In addition, it is unclear how one may, through an analysis analogous to that in Sect. 3, cast the above derivation in a more conventional form using the shape function and an "entropy of mixing" term.

Acknowledgements The author acknowledges helpful conversations with Prof. Roman F. Nalewajski and Prof. Robert G. Parr. Funding from NSERC, McMaster University, and the Canada Research Chairs is appreciated.

References

1. Hirshfeld FL (1977) Theor Chim Acc 44:129
2. Nalewajski RF (2003) J Phys Chem A 107:3792
3. De Proft F, Vivas-Reyes R, Peeters A, Van Alsenoy C, Geerlings P (2003) J Comp Chem 24:463
4. Roy RK (2003) J Phys Chem A 107:397
5. Nalewajski RF, Switka E (2002) Phys Chem Chem Phys 4:4952
6. Bultinck P, Langenaeker W, Lahorte P et al. (2002) J Phys Chem A 106:7895
7. De Proft F, Van Alsenoy C, Peeters A, Langenaeker W, Geerlings P (2002) J Comp Chem 23:1198
8. Ayers PW, Morrison RC, Roy RK (2002) J Chem Phys 116:8731
9. Olah J, Van Alsenoy C, Sannigrahi AB (2002) J Phys Chem A 106:3885
10. Ayers PW (2000) J Chem Phys 113:10886
11. Roy RK, Hirao K, Pal S (2000) J Chem Phys 113:1372
12. Roy RK, Pal S, Hirao K (1999) J Chem Phys 110:8236
13. Nalewajski RF, Broniatowska E (2003) Chem Phys Lett 376:33
14. Nalewajski RF (2003) Chem Phys Lett 372:28
15. Nalewajski RF (2002) Phys Chem Chem Phys 4:1710
16. Nalewajski RF, Switka E, Michalak A (2002) Int J Quantum Chem 87:198
17. Nalewajski RF, Parr RG (2001) J Phys Chem A 105:7391
18. Nalewajski RF, Loska R (2001) Theor Chem Acc 105:374
19. Nalewajski RF (2000) J Phys Chem A 104:11940
20. Davidson ER, Chakravorty S (1992) Theoretica Chimica Acta 83:319
21. Nalewajski RF (2003) Molecular Phys 101:2369
22. Nalewajski RF, Parr RG (2000) Proc Natl Acad Sci USA 97:8879
23. Kullback S (1997) Information theory and statistics. Dover, Mineola
24. Khinchin AI (1957) Mathematical foundations of information theory. Dover, New York

---

[2] Mathematically speaking, the perturbations should have compact support or, failing that, "essentially compact" support

25. Zucker R (1965) In: Abramowitz M, Stegun IA (eds) Handbook of mathematical functions. Dover, New York, p 65
26. Stuart A, Ord JK (1994) Kendall's Advanced theory of statistics, vol 1. Distribution theory. Halsted, New York
27. Abe S, Rajagopal AK (2003) Science 300:249
28. Vives E, Planes A (2002) Phys Rev Lett 8802:art-020601
29. Yamano T (2001) Phys Rev E 6304:art-046105
30. Rajagopal AK, Abe S (1999) Phys Rev Lett 83:1711
31. Tsallis C (1988) J Stat Phys 52:479
32. Ayers PW (2000) J Chem Phys 113:10886
33. Ayers PW (2000) Proc Natl Acad Sci USA 97:1959
34. Parr RG, Bartolotti LJ (1983) J Phys Chem 87:2810
35. Parr RG, Ayers PW, Nalewajski RF (2005) J Phys Chem A 109:3957
36. Sears SB, Parr RG, Dinur U (1980) Isr J Chem 19:165
37. Bader RFW (1990) Atoms in molecules: a quantum theory. Clarendon, Oxford
38. Bader RFW, Tal Y, Anderson SG, Nguyen-Dang TT (1980) Isr J Chem 19:8
39. Gelfand IM, Fomin SV (1991) Calculus of variations. Dover, Mineola