This paper investigates the relation between mathematics and minimalism and claims that most, if not all, of the principles of and constraints on syntactic movement advocated in minimalism can be deduced from mathematical considerations, meaning that such principles and constraints no longer need to be stipulated in syntactic theories and that they receive strong independent rationale from mathematics.

To achieve this goal, we first consider the nature of Merge (Chomsky 1995 et seq.). Let a, b, and c be either a lexical item or a phrase. Then, Merge takes two of them, say, b, and c, and forms a set \{b, c\}. Since Merge is recursive, this set can be an input to next Merge with a, yielding \{a, \{b, c\}\}. The set so formed has its own internal structure or a derivational history. In mathematical terms, we say it fails to satisfy the associativity law. Thus, \{a, \{b, c\}\} \neq \{{a, b}, c\}. This is illustrated in (1)

\[\{\text{black taxi} \text{ driver}\} = \{x: x \text{ is a driver of a black-colored taxi}\}\]

Let us note this point in (2) for the sake of reference.

In the domain of real numbers, the lack of associativity can be seen in division. E.g., \(10 \div (2 \div 2)=10, \text{while } (10 \div 2) \div 2=2.5\).

Putting (2) aside for the moment, let us now consider the distributive property (aka factorization by a common factor): \(ab + ac = a(b+c)\). In the domain of syntax, there is a likely candidate of operation that exhibits this property, namely, ATB movement, as illustrated in (3).

\[\text{[John bought a book] and [John read it]} \rightarrow \text{John [[bought a book] and [read it]]}\]

The ATB movement can be seen a process of (i) factoring out the common element John in the coordinated structure of the input sentence, and (ii) re-combining it with the coordinated verb phrases in the output. We identify the ATB transformation with the formula deformation: \(ab + ac = a(b+c)\). If this is correct, then it follows from the definition of the distributive property that syntax involves operations that correspond to addition and multiplication in mathematics. Formulating this in minimalist syntactic terms, we can hypothesize (4).

\[\text{There are two types of Merge; Addition Merge (AM) and Multiplication Merge (MM). For any element } a, b \text{ where } a, b \text{ are lexical items or phrases, AM } (a, b) = a+b, \text{ MM } (a, b) = a*b.\]

Applying this hypothesis to (3), it can be analyzed as in (5), where + stands for AM and * MM, and the square brackets [ ] are replaced by the parentheses ( ), and \(\rightarrow\) is by =. Furthermore, 1 is given as a multiplicative identity. Though it may seem like traces/copies but is different in that it does not have an index or represent a copy identity that makes it distinct from other occurrences of 1.

(5) \((\text{John}*\text{bought}*\text{a*book}) + (\text{John}*\text{read*it}) = \text{John}*(1*\text{bought}*\text{a*book} + 1*\text{read*it})\)

We call this analysis factorization approach (FA) and will generalize it to other types of movement.

As a first step of generalizing the FA, let us discuss DP movement. Consider (6), which illustrates an intermediate stage of the derivation of \(\text{John loves Mary}\), irrelevant details being omitted.

(6) \(T\varphi [_{\varphi} \text{John}\varphi \text{loves Mary}] \rightarrow [_{TP} \text{John}\varphi T\varphi [_{\varphi} <\text{John}\varphi> \text{loves Mary}]]\)

Accordingly to the standard analysis, \(T\varphi\) Agrees with \(\text{John}\varphi\) in \(\varphi\) and the latter moves into Spec-TP. In order to implement this analysis under FA, we take \(\varphi\) as a common factor between \(T\varphi\) and \(\text{John}\varphi\), and propose that movement of the latter can be regarded as a process of factoring out more than the common factor. Thus, under our theory, (6) can be analyzed as in (7).

(7) \(T\varphi + [_{\varphi} \text{John}\varphi \ast\text{loves}\ast\text{Mary}] = \text{John}\varphi\ast\left(\frac{T\varphi}{\text{John}\varphi} + 1 \ast\text{loves } \ast\text{Mary}\right)\)

In the left formula, in order to make factorization possible, we assume that \(T\varphi\) enters into derivation by AM and the rest by MM. Here, the square bracket and the vP label are not left out to keep the familiarity to the reader. In the right formula, \(\text{John}\varphi\) is factored out. Since it is more than \(\varphi\), say, some product of \(\varphi\), D and so on, \(T\varphi\) must be divided by \(\text{John}\varphi\) so as to make the two formulae equal. T. Though not shown in (7), reduction of fraction is possible. Thus, \(\frac{T\varphi}{\text{John}\varphi} = \frac{T\varphi}{\text{John}\varphi}\). We take this as feature-
checking/deletion/valuation.

Let us now consider how movement should be constrained under the FA, in light of (2). Given that, it should be impossible to move an element of a set inside another set (or more correctly, to move a factor of a product which in turn is a factor of another product). This is illustrated in (9), where the formula is deformed, with c as a common factor, but the equality does NOT hold, as a*(b*c) is lost.

\[(9) \ c + a*(b*c) \neq c*\{1+a*(b*1)\}\]

Notice that in case of DP movement observed in (7), factorization is possible because the common factor is that factor which underwent MM last. That is, unlike (9), it has the general (simplified) form in (10).

\[(10) a + a*(b*c) = a*\{1+1*(b*c)\}\]

In syntactic terms, this means that only the edge element can move, a constraint known as the Phase Impenetrability Condition (PIC) proposed by Chomsky (2001, 2004). Thus, we have derived (a very strong version of) the PIC from the combination of (2) and the factorization analysis of movement.

So far so good. However, now we need to face a problem posed by wh-movement. To understand the problem, consider the stage of a derivation in (11a), where whoWH moves to outer Spec-VP so as to move to Spec-CP in the next higher phase. The next step is Merge of CWH and Tφ, as in (11b). Then, CWH first moves whoWH to its Spec, as in (11c), followed by the subject DP movement triggered by Tφ, as in (11d).

\[(11) a, [\_p Johnφ \_v [VP love whoWH]] \rightarrow [\_p whoWH [\_v Johnφ \_v [VP love <whoWH>]]]
\]

Consider how movement should be constrained under the FA, in light of (2). Given that, it should be impossible to move an element of a set inside another set (or more correctly, to move a factor of a product which in turn is a factor of another product). This is illustrated in (9), where the formula is deformed, with c as a common factor, but the equality does NOT hold, as a*(b*c) is lost.

\[(10) a + a*(b*c) = a*\{1+1*(b*c)\}\]

In syntactic terms, this means that only the edge element can move, a constraint known as the Phase Impenetrability Condition (PIC) proposed by Chomsky (2001, 2004). Thus, we have derived (a very strong version of) the PIC from the combination of (2) and the factorization analysis of movement.

So far so good. However, now we need to face a problem posed by wh-movement. To understand the problem, consider the stage of a derivation in (11a), where whoWH moves to outer Spec-VP so as to move to Spec-CP in the next higher phase. The next step is Merge of CWH and Tφ, as in (11b). Then, CWH first moves whoWH to its Spec, as in (11c), followed by the subject DP movement triggered by Tφ, as in (11d).

\[(11) a, [\_p Johnφ \_v [VP love whoWH]] \rightarrow [\_p whoWH [\_v Johnφ \_v [VP love <whoWH>]]]
\]

This derivation entails countercyclicity in that wh-movement precedes DP-movement, which is required because the wh-phrase is closer to T than the subject at (11b) stage, which in turn is necessitated by the PIC at (11a) stage. The reasoning sounds rather theory-internal. Is there any independent support for it?

Now that the PIC is deduced from the proposed model, can we deduce the countercyclicity? No, wait! How should we deal with wh-movement in the first place? Our PIC is stronger than Chomsky’s because it permits only an edge element to move, which prohibits even the intermediate wh-movement in (11a). How should we implement wh-movement? Here a solution comes from mathematics. Consider how multiplication of constants behaves when they appear in division calculation. As an illustration of the lack of associativity, we saw that \(10 ÷ (2 ÷ 2) = 10\), but \((10 ÷ 2) ÷ 2 = 2.5\). However, if we multiply by \(-1\) one of the numbers in \(10 ÷ (2 ÷ 2)\), the results are the same whichever number it is multiplied with. Thus, \(10∗\{-2 ÷(2 \times -1)\} = 10∗\{2 ÷ (2 \times -1)\ = \{10 \times (-1)\} ÷ (2 ÷ 2) = -10\). This shows that it is possible to combine the mathematical objects and operations that satisfy the associativity law with those that do not.

So we propose that wh-phrases are syntactic constants that satisfy the associativity law, having the “freedom of movement.” Then, wh-movement can be analyzed as follows. Here let us recycle (11) for illustration. At (11a) stage, the wh-phrase does not have to move because it is a constant that satisfies the associativity law. After CWH and Tφ enter into derivation by AM, structure (12) obtains. From here, wh-phrase is first factored out, yielding A, followed by factoring out of Johnφ, as illustrated in B.

\[(12) C_{WH} + Tφ + [\_p Johnφ \_v [VP love whoWH]]
\]

Notice that factorization of the wh-phrase must precede that of the subject. Otherwise, (13) ensues.

\[(13) C_{WH} + Johnφ ∗ \left( \frac{Tφ}{Johnφ} + 1 ∗ v \_v love \_v whoWH \right)\]

From (13), it is impossible to factor out the wh-phrase even if it can be dealt with as a constant. This shows that the countercyclicity in syntax has a mathematical rationale, which frees syntax from all the stipulations made to implement it such as feature-inheritance and so on (Chomsky 2005). Finally, the island phenomena can all be attributed to ill-formed factorization, just like in (13), i.e., wh-movement intended from a structure which is already factorized (wh-islands and CNPC) or from a coordinated structure where there is no common factor (CSC and adjunct islands). References Chomsky (1995) The Minimalist Program, (2001) Derivation by phase, (2005) Three factors in language design, LI 36.