On Linear Inference

Frank Pfennnig
Draft of February 2, 2008

Inference. When we write an inference rule

$$\frac{\mathsf{even}(x)\ \mathit{true}}{\mathsf{even}(\mathsf{s}(\mathsf{s}(x)))\ \mathit{true}}$$

what do we mean? To discuss this, some terminology: we say "$\mathsf{even}(t)$" is a *proposition* and "$\mathsf{even}(t)$ *true*" is a *judgment*. Following Martin-Löf, a judgment is an *object of knowledge*. We obtain knowledge by making inferences from judgments we already know. We can then read the rule above as

> *If we know that t is even for a term t, we may conclude (and thereby know) that $\mathsf{s}(\mathsf{s}(t))$ is also even.*

The process of inference is therefore one by which we gain knowledge. We may start with the knowledge that $\mathsf{0}$ is even, then we gain the information that $\mathsf{s}(\mathsf{s}(\mathsf{0}))$ is even, and next we learn that $\mathsf{s}(\mathsf{s}(\mathsf{s}(\mathsf{s}(\mathsf{0}))))$ is even, and so on. This process is clearly monotonic: we gain knowledge, but we never forget, at least in the idealized world of mathematics.

Persistent and Ephemeral Truth. It has long been argued by philosophers that truth is not universal, but depends on the world in which we consider a proposition. This could depend on time ("*It is raining now, but it did not rain yesterday.*"), place ("*it is hot in the Sahara but cold at the North Pole.*") or simply the state of a circumscribed system ("*The white king is on square e1.*"). Studying logical inference in these situations is the domain of temporal, spatial, or linear logic, respectively. In this note we discuss the latter.

We start by distinguishing two basic judgment forms, generalizing truth: proposition A is *persistently true* (A *pers*) and A is *ephemerally true* (A *eph*). Persistent truths are like the mathematical truths we have discussed earlier: the number 2 is even, and this does not change. Ephemeral truths are those that subject to change, those that depend on the current state. For example, after we make a move, the white king may no longer be on square e1 and so this proposition is ephemeral.

What is the nature of *inference* in the setting where we have persistent and ephemeral truths? As an example, we consider the familiar blocks world planning problems, with propositions $\mathsf{on}(x, y)$ (if x is on y), $\mathsf{clear}(x)$ (if x the top of x is clear and x can therefore be picked up), free (the robot hand is free) and $\mathsf{holds}(x)$ (the robot hand holds block x). We describe the possibility of picking up a block with the following rule:

$$\frac{\begin{array}{l}\mathsf{on}(x, y)\ \mathit{eph} \\ \mathsf{clear}(x)\ \mathit{eph} \\ \mathsf{free}\ \mathit{eph}\end{array}}{\begin{array}{l}\mathsf{holds}(x)\ \mathit{eph} \\ \mathsf{clear}(y)\ \mathit{eph}\end{array}}\ \mathsf{pickup}$$

All judgments here concern ephemeral truth, and the rule can be read as follows:

> *If we are in a state where x is on y, x is clear, and the hand is free, we can reach a new state where the hand holds x and y is clear.*

Performing this inference means that the ephemeral propositions in the premiss are no longer true in the new state, and the propositions in the conclusion are now true in the new state.

An inference rule allows us to perform an inference, but does not force us to do so. In the case of persistent (mathematical) truth, this means we may never learn that 1592847498 is even. In the case of emphemeral truth, this means we may never pick up a given block, even if the rule pickup would permit us to do so. This is more important in this new setting because inferences may be irreversible, so making an inference may constitute a real commitment. If all truths are persistent (and hence inference is monotonic) we can always decide to postpone application of a rule since it can be applied at any future point in our deduction process.

New Forms of Rules. There is another new phenomenon, which is that inference rule must be allowed to have not only multiple premisses, but also multiple conclusions. The example above demonstrates this: we cannot break pickup into two rules each with a single conclusion, not only because the premisses are consumed during the inference, but also because both conclusions must appear in the state during the same atomic step. In our model of blocks world, there is no intermediate state where y is clear but we do not yet hold x.

Similarly, it must be permissible for an inference rule to have no conclusion. For example, if we had the possibility to discard a block that is on the table, the rule might read

$$\frac{\begin{array}{l}\mathsf{clear}(x)\ \mathit{eph} \\ \mathsf{on}(x, \mathsf{table})\ \mathit{eph}\end{array}}{}\ \mathsf{discard}$$

Mixing Persistent and Ephemeral Propositions. New considerations arise when persistent and ephemeral propositions are combined. Consider that blocks may either be small or large, and that the hand can only pick up small blocks. Since the size of the blocks does not change, $\mathsf{small}(x)$ is persistent.

$$\frac{\begin{array}{l}\mathsf{small}(x)\ \mathit{pers}\\ \mathsf{on}(x,y)\ \mathit{eph}\\ \mathsf{clear}(x)\ \mathit{eph}\\ \mathsf{free}\ \mathit{eph}\end{array}}{\begin{array}{l}\mathsf{holds}(x)\ \mathit{eph}\\ \mathsf{clear}(y)\ \mathit{eph}\end{array}}\ \mathsf{pickup}$$

The persistent nature of $\mathsf{small}(x)$ means that when a block x is picked up by using this inference, the block remains small. Simple.

A slightly more subtle observation is that a persistently true proposition can be used to satisfy the premiss of a rule $A\ \mathit{eph}$. We can think of the premiss $A\ \mathit{eph}$ as saying that A must be true in the state before the inference. Certainly,if A is true in all states (that is, persistent) then it is true in any particular state where we might need this fact. For example, we could represent a robot with more hands than blocks with the persistent fact free. Now, any time during inference we need to know that a hand is free, we can use this without destroying it.

Multiplicity. One final observation is that ephemeral truth can have a multiplicity. For example, we might have three nickels, represented by three judgments

$$\mathsf{nickel}\ \mathit{eph}, \mathsf{nickel}\ \mathit{eph}, \mathsf{nickel}\ \mathit{eph}.$$

The rule for changes two nickels into a dime would be

$$\frac{\begin{array}{c}\mathsf{nickel}\ \mathit{eph}\\ \mathsf{nickel}\ \mathit{eph}\end{array}}{\mathsf{dime}\ \mathit{eph}}$$

After applying this rule in the state above we have, as expected,

$$\mathsf{nickel}\ \mathit{eph}, \mathsf{dime}\ \mathit{eph}.$$

Which two of the three nickels we changed into a dime remains ambiguous, but can be made precise if we label all components of a state with a unique label, like a variable labeling each assumption in a deduction under the Curry-Howard isomorphism.

In the whole discussion above, we only used atomic propositions, judgments about them, and rules of inference. If we systematically construct a logic from these considerations by internalizing the various notions, we obtain the judgmental formulation of intuitionistic linear logic by Chang et al. Relating propositions in linear logic back to the above formulation works only in a Horn-like fragment, but is otherwise straightforward using the tools of focusing (see, for example, the course notes on *Logic Programming*). We will do so now.

A Horn-Like Fragment of Linear Logic. The fragment of linear logic we consider allows us to express all of the above, now within a logic rather than entirely at the level of judgments. This fragment is remiciscent of Horn logic, but mixes ephemeral and persistent propositions.

Atomic propositions	P		
Basic propositions	Q	$::=$	$P \mid !P$
State propositions	S	$::=$	$Q \mid \mathbf{1} \mid S \otimes S$
State transitions	R	$::=$	$S \multimap S \mid \forall x.\, R$
Persistent hypotheses	Γ	$::=$	$\cdot \mid \Gamma, R\ \mathit{pers} \mid \Gamma, P\ \mathit{pers}$
Ephemeral hypotheses	Δ	$::=$	$\cdot \mid \Delta, R\ \mathit{eph} \mid \Delta, P\ \mathit{eph}$

The sequents defining the meaning of the connectives on this fragment have the form

$$\Gamma; \Delta \vdash S\ \mathit{eph}$$

where Γ consists of persistent assumptions, Δ consists of ephemeral assumptions, and S is the conclusion, a proposition describing a state. The inference rules are the standard ones from the judgmental formulation of linear logic, so we do not give them here.

The connection between reasoning in linear logic and the reconstruction of persistent and ephemeral truth can be explored in both directions. In one direction, alluded to above, we can internalize the judgment forms and obtain linear logic with its full range of connective. In the other direction, we can apply focusing to show that the inference rule formulation is indeed sound and complete with respect to small step reasoning in linear logic. In order to do this, we first present the focusing rules for linear logic on the fragment above. All atomic propositions are considered *positive*, as are $\otimes$ and $\mathbf{1}$, while universal quantification and linear implication are negative. We further restrict Δ to consist only of (positive) atoms. We obtain the rules below. Since the judgments *eph* and *pers* can be inferred from the position in the sequent, we elide them below to shorten the rules.

Focusing. First, from a neutral sequent $\Gamma; \Delta \vdash S$ we can focus either on an assumption R in Γ or on the right-hand side S. We cannot focus on something in Δ because it contains only positive atoms.

$$\frac{\Gamma; \Delta; [R] \vdash S \quad R \in \Gamma}{\Gamma; \Delta \vdash S}\ \mathsf{focus}L! \qquad \frac{\Gamma; \Delta \vdash [S]}{\Gamma; \Delta \vdash S}\ \mathsf{focus}R$$

Once we are focused on the left, we decompose the negative connectives, universal quantification and implication.

$$\frac{\Gamma; \Delta; S(t) \vdash S}{\Gamma; \Delta; [\forall x.\, R(x)] \vdash S}\ \forall L \qquad \frac{\Gamma; \Delta_1 \vdash [S_1] \quad \Gamma; \Delta_2; [S_2] \vdash S}{\Gamma; (\Delta_1, \Delta_2); [S_1 \multimap S_2] \vdash S}\ {\multimap}L$$

When the left focus formula becomes positive, that is, a state formula S, we transition to a left inversion phase.

$$\frac{\Gamma; \Delta; S' \vdash S}{\Gamma; \Delta; [S'] \vdash S}\ \mathsf{blur}L$$

The left inversion phase will decompose $\otimes$ and $\mathbf{1}$ and move atomic formulas into either Γ (if they are persistent) or Δ (if they are ephemeral).

$$\frac{\Gamma; \Delta; \Psi \vdash S}{\Gamma; \Delta; \Psi, \mathbf{1} \vdash S}\ \mathbf{1}L \qquad \frac{\Gamma; \Delta; \Psi, S_1, S_2 \vdash S}{\Gamma; \Delta; \Psi, S_1 \otimes S_2 \vdash S}\ \otimes L$$

$$\frac{(\Gamma, P); \Delta; \Psi \vdash S}{\Gamma; \Delta; \Psi, !P \vdash S}\ !L \qquad \frac{\Gamma; (\Delta, P); \Psi \vdash S}{\Gamma; \Delta; \Psi, P \vdash S}\ \mathsf{atom}L$$

When all negative connectives have been decomposed, we transition back to a neutral sequent, ready to start another focusing phase.

$$\frac{\Gamma; \Delta \vdash S}{\Gamma; \Delta; \cdot \vdash S}\ \mathsf{deactivate}L$$

When we focus on the right, we decompose the state formula until we reach an basic proposition. This must either be directly in the context (if it is emphemeral), or we lose focus and the underlying persistent propositions must follow from the persistent hypotheses alone. This final sequent is also neutral.

$$\frac{}{\Gamma; \cdot \vdash [\mathbf{1}]}\ \mathbf{1}R \qquad \frac{\Gamma; \Delta_1 \vdash [S_1] \quad \Gamma; \Delta_2 \vdash [S_2]}{\Gamma; \Delta_1, \Delta_2 \vdash [S_1 \otimes S_2]}\ \otimes R$$

$$\frac{}{\Gamma; P \vdash [P]}\ \mathsf{atom}R \qquad \frac{\Gamma; \cdot \vdash P}{\Gamma; \cdot \vdash [!P]}\ !R$$

A Prototypical Example. As a prototypical example we consider the rule

$$\frac{\begin{array}{l} A\ \mathit{pers} \\ B\ \mathit{eph} \end{array}}{\begin{array}{l} C\ \mathit{eph} \\ D\ \mathit{pers} \end{array}}$$

where A, B, C, D, E are all atomic propositions. In linear logic, this rule can be expressed as

$$!A \otimes B \multimap C \otimes !D.$$

Call this proposition R_0 and add it to Γ, since the inference rules themselves are persistent in the paradigm we are considering.

What does it mean to focus on R_o, using the $\textsf{focus}L!$ rule? Because a focused derivations are quite restricted, the shape of the derivation is mostly determined. At first, we just unwind R_0, leaving contexts and also the right-hand side indeterminate. We stop at initial sequents or neutral sequents.

$$\dfrac{\dfrac{\dfrac{\dfrac{\_;\_ \vdash A}{\_;\_ \vdash [!A]}\;!R \qquad \dfrac{}{\_;\_ \vdash [B]}\;\textsf{atom}R}{\_;\_ \vdash [!A \otimes B]}\;\otimes R \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\_,D;\_,C \vdash \_}{\_,D;\_,C;\cdot \vdash \_}\;\textsf{deactivate}L}{\_,D;\_;C \vdash \_}\;\textsf{atom}L}{\_;\_;C,!D \vdash \_}\;!L}{\_;\_;C \otimes !D \vdash \_}\;\otimes L}{\_;\_;[C \otimes !D] \vdash \_}\;\textsf{blur}L}{\_;\_;[R_0] \vdash \_}\;\multimap L \qquad R_0 \in \_}{\_;\_ \vdash \_}\;\textsf{focus}L!$$

Filling in contexts, we notice first that the use of $\textsf{atom}R$ requires B to be a linear hypothesis. We further notice that the use of $!R$ requires there to be no ephemeral hypothesis in the subgoal proving A. Propagating this information yields:

$$\dfrac{\dfrac{\dfrac{\dfrac{\_;\cdot \vdash A}{\_;\cdot \vdash [!A]}\;!R \qquad \dfrac{}{\_;B \vdash [B]}\;\textsf{atom}R}{\_;B \vdash [!A \otimes B]}\;\otimes R \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\_,D;\_,C \vdash \_}{\_,D;\_,C;\cdot \vdash \_}\;\textsf{deactivate}L}{\_,D;\_;C \vdash \_}\;\textsf{atom}L}{\_;\_;C,!D \vdash \_}\;!L}{\_;\_;C \otimes !D \vdash \_}\;\otimes L}{\_;\_;[C \otimes !D] \vdash \_}\;\textsf{blur}L}{\_;\_,B;[R_0] \vdash \_}\;\multimap L \qquad R_0 \in \_}{\_;\_,B \vdash \_}\;\textsf{focus}L!$$

Next, we fill in the remaining holes parametrically and summarize the new derived rule where premisses and conclusion are just neutral sequents.

$$\dfrac{\Gamma;\cdot \vdash A \qquad (\Gamma,D);(\Delta,C) \vdash S \qquad R_0 \in \Gamma}{\Gamma;(\Delta,B) \vdash S}$$

This gives a strong hint on how to read the view of inference from the introduction as a focused proof search strategy in linear logic. We think of Γ as the persistently true propositions, Δ the propositions ephemerally true in the current state. We read the rule from the conclusion to the premise, as a step in the search for a proof.

> *If we know the ephemeral proposition B, and the persistent proposition A follows from the all known persistent propositions Γ, then*

we can reach a state where D is persistently true and C is a new ephemeral truth while rescinding B.

This reading does not quite coincide with our notion of linear inference discussed first in this note. The mismatch is that arbitrarily complex reasoning can be used to establish A, and this reasoning is to the side of our main inference activity. This is clearly undesirable, because it means that the question if a rule can be applied in a given state may be undecidable simply because of its persistent premisses.

The analysis clearly shows that some additional restrictions are necessary to bring inference with ephemeral and persistent truths in concordance with focused linear logic.

Saturation. In preparation for the final step, we return to the setting where all judgments are persistent truths. If we want to know if an atomic proposition P follows from a known collection of atomic propositions and a set of rules, we can simply keep applying rules when their premisses are satisfied. If we do this in a fair manner (every rule that can be applied will eventually be applied), this process is complete and if P is indeed true, it must eventually be derived.

We say the set of known atomic propositions is *saturated* if applying any inference rule does not add any new knowledge. If our known facts are saturated, then P follows precisely if we have already deduced it, so the test for P can be reduced by a lookup.

In many cases we can present rules in such a way that the inference process must always saturate. We then write $\mathsf{Clo}(\Gamma)$ for the closure of Γ under the rules of inference. Moreoever, we can often determine the complexity of the saturations process. It is evident from the foregoing that in such theories we have $\Gamma \vdash P$ if and only of $P \in \mathsf{Clo}(\Gamma)$.

Separation. Unfortunately, in the presence of ephemeral hypotheses and conclusions, saturation is not well-defined even if we start with no ephemeral truth. For example, the rules

$$\frac{a\ \mathit{pers}}{b\ \mathit{eph}} \qquad \frac{b\ \mathit{eph}}{}$$

can generate and consume an arbitrary number of ephemeral b. We additional restrictions to so that inference with persistent and ephemeral propositions truly coincides with reasoning in linear logic. The analysis of focused reasoning suggests that we need a sufficient condition so that

$$\Gamma; \cdot \vdash P \quad \text{iff} \quad P \in \mathsf{Clo}(\Gamma)$$

If we can achieve this, then the rule on the left, read in the forward direction (from premisses to conclusion), corresponds to the focusing rule on the right, read in the backward direction (from conclusion to premisses).

$$\frac{\begin{array}{c} A \; pers \\ B \; eph \end{array}}{\begin{array}{c} C \; eph \\ D \; pers \end{array}} \qquad \frac{A \in \mathsf{Clo}(\Gamma) \quad (\Gamma, D); (\Delta, C) \vdash S}{\Gamma; (\Delta, B) \vdash S}$$

We would probably want to arrange things so that Γ is saturated with respect to persistent propositions, so that the linear logic derived rule further simplifies, just checking that $A \in \Gamma$.

Two straightforward conditions amount to a form of separation between persistent and ephemeral reasoning

1. Every predicate symbol occurs only persistently or ephemerally. This means no persistent proposition can realize an ephemeral premiss of a rule.

2. Every rule with an ephemeral conclusion also has at least one ephemeral premiss. This prevents us from circumventing the first restriction by introducing a new (renamed) propositions.

We conjecture that these are sufficient to guarantee that, on our fragment, $\Gamma; \cdot \vdash P$ iff $P \in \mathsf{Clo}(\Gamma)$.

Range Restriction. So far we have concentrated on the propositional aspects of inference, but the terms and parameters in the inference rules also present some challenges. One difficulty can be see in a rule such as

$$\frac{}{\mathsf{positive}(\mathsf{s}(x))}$$

which expresses that any number of the form $\mathsf{s}(t)$ for some t is positive. Since there is no premiss, we can apply this inference for any t in place of x, but this yields infinitely many different conclusions.

One way to resolve this to allow parametric truths to be asserted, rather than just ground truths. This leads to what is traditionally called *resolution*, where any clause is parametric in all of its free variables.

Saturation and complexity are easier to analyze if we restrict ourselves only to ground truths. In that case, we may restrict rules to be *range restricted* which means that every variable in the conclusion of the rule must also appear in at least one premise. If the premisses are matched against ground truths (whether ephemeral or persistent does not matter), then all variables in a rule will be instantiated to ground terms, and the conclusion will also again be ground.

This level of choice is not represent in the focusing system in the form above, but can easily be accomodated by replacing guessed terms (in the $\forall L$ rule) by meta-variables subject to unification (in the general case) or matching (in the range-restricted case). A blueprint for this kind of generalization can be found in the Logic Programming course notes.